

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

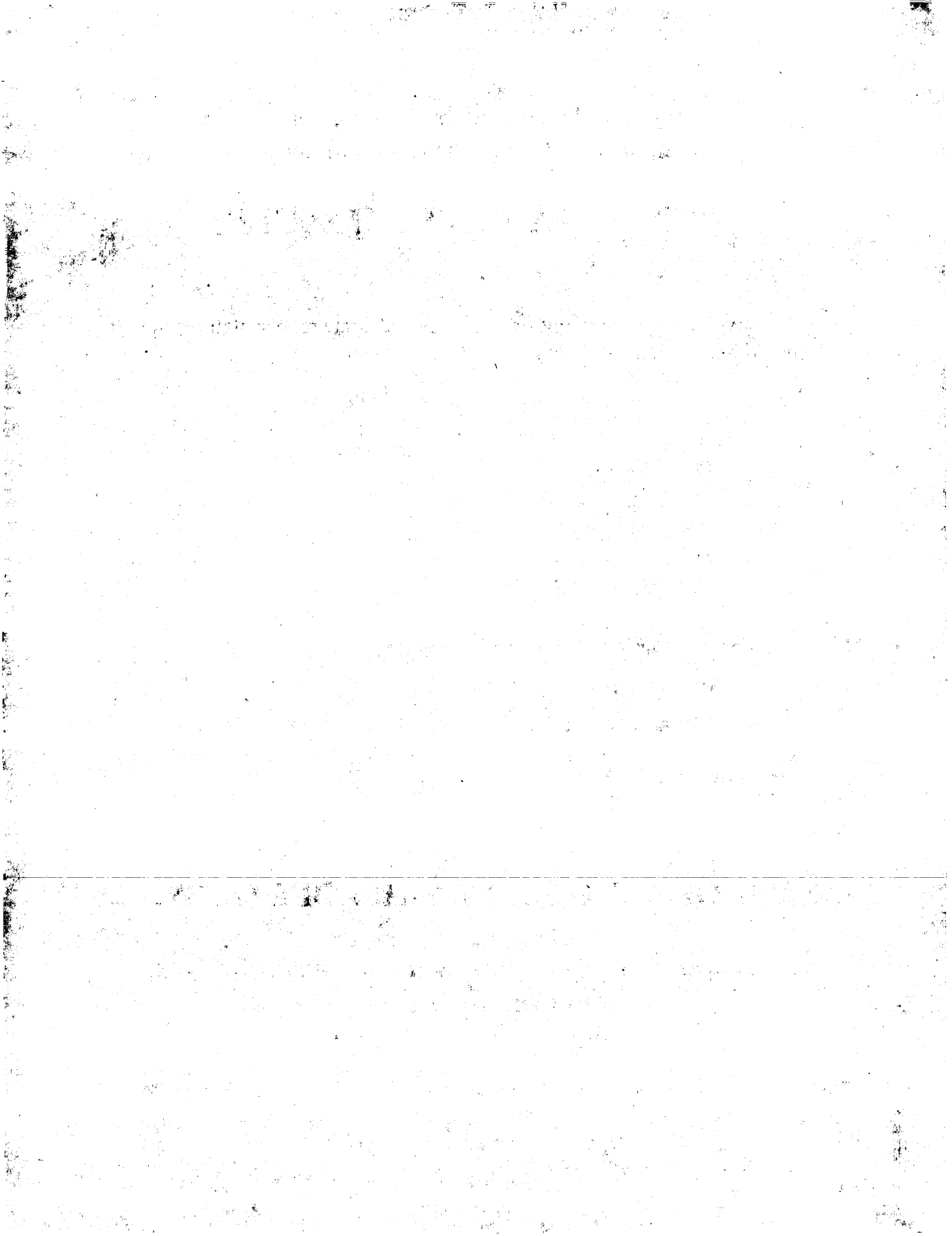
Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**





INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁵ : C12N 15/12, 15/54, 9/12 // C12Q 1/68, C12N 15/11</p>	A1	<p>(11) International Publication Number: WO 93/15201</p> <p>(43) International Publication Date: 5 August 1993 (05.08.93)</p>		
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top; padding: 5px;"> <p>(21) International Application Number: PCT/US93/00586</p> <p>(22) International Filing Date: 22 January 1993 (22.01.93)</p> <p>(30) Priority data: 826,935 22 January 1992 (22.01.92) US</p> <p>(60) Parent Application or Grant (63) Related by Continuation US 826,935 (CIP) Filed on 22 January 1992 (22.01.92)</p> <p>(71) Applicant (for all designated States except US): NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).</p> </td> <td style="width: 50%; vertical-align: top; padding: 5px;"> <p>(72) Inventors; and (75) Inventors/Applicants (for US only) : AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US).</p> <p>(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).</p> <p>(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p> </td> </tr> </table>			<p>(21) International Application Number: PCT/US93/00586</p> <p>(22) International Filing Date: 22 January 1993 (22.01.93)</p> <p>(30) Priority data: 826,935 22 January 1992 (22.01.92) US</p> <p>(60) Parent Application or Grant (63) Related by Continuation US 826,935 (CIP) Filed on 22 January 1992 (22.01.92)</p> <p>(71) Applicant (for all designated States except US): NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).</p>	<p>(72) Inventors; and (75) Inventors/Applicants (for US only) : AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US).</p> <p>(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).</p> <p>(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>
<p>(21) International Application Number: PCT/US93/00586</p> <p>(22) International Filing Date: 22 January 1993 (22.01.93)</p> <p>(30) Priority data: 826,935 22 January 1992 (22.01.92) US</p> <p>(60) Parent Application or Grant (63) Related by Continuation US 826,935 (CIP) Filed on 22 January 1992 (22.01.92)</p> <p>(71) Applicant (for all designated States except US): NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).</p>	<p>(72) Inventors; and (75) Inventors/Applicants (for US only) : AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US).</p> <p>(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).</p> <p>(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>			
<p>(54) Title: NOVEL PROTEIN TYROSINE KINASES</p>				
<p>(57) Abstract</p> <p>The identification and isolation of novel protein tyrosine kinase genes present on human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins. RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.</p>				

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CC	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

NOVEL PROTEIN TYROSINE KINASESDescriptionBackground of the Invention

Transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases are enzymes that catalyze this process. Moreover, many act as growth factor receptors.

Summary of the Invention

10 The present invention relates to novel protein tyrosine kinase genes present in human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of
15 use therefor.

 The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology with
20 previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. The pTK genes have been shown to be present in both megakaryocytic and lymphocytic cells.

 The pTK genes of the present invention show
25 significant sequence homology with members of the c-kit subgroup of growth factor receptors with protein tyrosine kinase activity. The c-kit subgroup of receptor tyrosine kinases catalyze the phosphorylation of exogenous substrates, as well as tyrosine residues within their own
30 polypeptide chains. (Ullrich, A. and Schlessinger, J., Cell, 61:203 (1990)). Members of the c-kit subgroup include FLT/FLK (Fetal Liver Kinase), FGF (Fibroblast

-2-

Growth Factor Receptor) and NGF (Nerve Growth Factor Receptor).

In particular, fourteen pTK genes have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor tyrosine kinase) were identified in megakaryocytic cells. Five pTK genes, referred to as LpTKs, were identified in lymphocytic cells and have been shown to be present in megakaryocytes as well. One pTK gene, referred to as HpTKs, was identified in human hepatoma cells. Six pTK genes, referred to as bpTK genes, found in human brain tissue.

SAL-S1 is related to the FLT/FLK family of pTKs. SAL-D4 is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3) is related to the NGF receptor family of pTKs.

The pTK genes, which are the subject of the present invention, were identified using two sets of degenerative oligonucleotide primers: a first set which amplifies all pTK DNA segments (SEQ ID NOS:1-2), and a second set which amplifies highly conserved sequences present in the catalytic domain of the c-kit subgroup of pTKs (SEQ ID NOS:3-4). The pTK genes identified in this manner are described below.

SAL-S1 is expressed in several megakaryocytic cell lines, but not in erythroid cell lines. The nucleotide sequence of SAL-S1 was obtained, revealing a sequence containing 158 base pairs. (SEQ ID NO:5). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:6) which exhibited significant sequence homology with known protein tyrosine kinases of the FLT/FLK family. The full

-3-

length gene sequence (SEQ ID NO: 17) contains 6827 b.p. and the deduced amino acid sequence (SEQ ID NO: 18) contains 349 residues.

SAL-D4, also expressed in megakaryocytic cells, is a DNA fragment containing the nucleotide sequence of 141 base pairs. (SEQ ID NO:7). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:8) which exhibited significant sequence homology with known protein tyrosine kinases of the FGF receptor family.

10 The LpTKs, including LpTK 2, LpTK 3, LpTK 4, and LpTK 13 and LpTK 25, are expressed in lymphocytic cells, as well as megakaryocytic cells. The nucleotide sequence (151 base pairs) of the LpTK 3 gene was obtained (SEQ ID NO:11), and exhibited significant homology with known
15 protein tyrosine kinases of the NGF receptor family. The nucleotide sequences of the LpTK 2, LpTK 4, and LpTK 13 genes contained 149 base pairs (SEQ ID NO:9), 137 base pairs (SEQ ID NO:13), and 211 base pairs (SEQ ID NO:15) respectively. LpTK 25 has a nucleotide sequence of 3120
20 b.p. (SEQ ID NO: 22). A full length gene sequence has been obtained for LpTK 2 (SEQ ID NO: 19) which contains 7606 b.p. Additional sequencing of LpTK 4 revealed a sequence of 404 b.p. (SEQ ID NO: 21).

The HpTK 5 gene, expressed in human hepatoma cells,
25 has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). Nucleotide sequences of the bpTK's, including bpTK 1, bpTK 2, bpTK 3, bpTK 4, bpTK 5 and bpTK 7 are expressed in human brain tissue encode proteins having the amino acid sequences of SEQ ID NOS: 25-30 respectively.

30 Thus the present invention includes DNA isolated from a human megakaryocytic cell line, which hybridizes to a

-4-

DNA fragment which hybridizes to DNA encoding an amino acid sequence which is highly conserved in the catalytic domain of protein tyrosine kinases of the c-kit subgroup.

The present invention also includes the proteins
5 encoded by the pTK genes identified as described herein, which exhibit significant sequence homology with members of the c-kit subgroup of pTKs (i.e. FLT/FLK (SAL-S1), FGF receptor (SAL-D4) or NGF receptor (LpTKS)) as well as the proteins encoded by HpTK 5 and the bpTKs. The present
10 invention also includes SAL-S1, SAL-D4, and LpTK, HpTK and bpTK homologues or equivalents (i.e., proteins which have amino acid sequences substantially similar, but not identical, to that of SAL-S1, SAL-D4, the LpTKs HpTK and the bpTKs, which exhibit tyrosine kinase activity.) This
15 invention further includes peptides (SAL-S1, SAL-D4, LpTK, HpTK and bpTK fragments) which retain tyrosine kinase activity, yet are less than the entire SAL-S1, SAL-D4, LpTK, HpTK and bpTK sequences), monoclonal and polyclonal antibodies specific for SAL-S1, SAL-D4, the LpTKs, HpTK
20 and the bpTKs, and uses for the SAL-S1, SAL-D4, the LpTK, HpTK and the bpTK nucleic acid sequences and SAL-S1, SAL-D4, LpTK, HpTK and bpTK equivalents.

The present invention further includes nucleic acid sequences which hybridize with DNA or RNA encoding the
25 proteins described herein, which exhibit significant sequence homology with the FLT/FLK, FGF receptor or NGF receptor family of protein tyrosine kinases contained within the c-kit subgroup. Such nucleic acid sequences are useful as probes to identify pTK genes in other
30 vertebrates, particularly mammals, and in other cell types. They can also be used as anti-sense

-5-

oligonucleotides to inhibit protein tyrosine kinase activity, both in vitro and in vivo.

The SAL-S1, SAL-D4, LpTK, HpTK and bpTK, tyrosine kinases of the present invention can be used as target
5 proteins in conjunction with the development of drugs and therapeutics to modulate cell growth, differentiation and other metabolic functions. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK proteins can be used as agonists or antagonists to other tyrosine kinases. The SAL-S1, SAL-D4, LpTK, HpTK or
10 bpTK tyrosine kinases can also be instrumental in the modulation of megakaryocyte and/or platelet adhesion interactions.

In addition, the SAL-S1, SAL-D4, LpTK, HpTK and bpTK tyrosine kinases can be used in screening assays to detect
15 cellular growth and/or differentiation factors. Using standard laboratory techniques, the ligands of the pTKs of the present invention can be identified. Once identified, assays can be designed to detect these ligands present endogenously, within cells, as well as exogenously, in
20 extra cellular fluids. Assays can also be designed as diagnostic aids to detect these ligands in body fluids such as blood and urine.

Brief Description of the Drawings

Figure 1 depicts the nucleotide sequence of SAL-S1
25 (SEQ ID NO: 5) and the deduced amino acid sequence (SEQ ID NO:6).

Figure 2 depicts the nucleotide sequence of SAL-D4 (SEQ ID NO:7) and its deduced amino acid sequence (SEQ ID NO:8).

-6-

Figure 3A depicts the nucleotide sequence (SEQ ID NO:9) and its deduced amino acid sequence (SEQ ID NO:10) for LpTK 2.

Figure 3B depicts the nucleotide sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) for LpTK 3.

Figure 3C depicts the nucleotide sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) for LpTK 4.

10 Figure 3D depicts the nucleotide sequence (SEQ ID NO:15) and its deduced amino acid sequence (SEQ ID NO:16) for the LpTK 13.

Figure 4A-4J depicts the full-length nucleotide sequence (SEQ ID NO: 17) and its deduced amino acid sequence (SEQ ID NO: 18) for SAL-S1.

Figure 5A-5J depicts the full length nucleotide sequence (SEQ ID NO: 19) and the deduced amino acid sequence (SEQ ID NO: 20) for LpTK2.

20 Figure 6 depicts the partial nucleotide sequence (SEQ ID NO: 21) for LpTK4.

Figure 7A-7D depicts the full length nucleotide sequence (SEQ ID NO: 22) for LpTK25.

Figure 8A-8F depicts the full length nucleotide sequence (SEQ ID NO: 23) and the deduced amino acid sequence (SEQ ID NO: 24) for HpTK5.

Figure 9 depicts the amino acid sequence (SEQ ID NO: 25) of bpTK1.

Figure 10 depicts the amino acid sequence (SEQ ID NO: 26) of bpTK2.

30 Figure 11 depicts the amino acid sequence (SEQ ID NO: 27) of bpTK3.

-7-

Figure 12 depicts the amino acid sequence (SEQ ID NO: 28) of bpTK4.

Figure 13 depicts the amino acid sequence (SEQ ID NO: 29) of bpTK5.

5 Figure 14 depicts the amino acid sequence (SEQ ID NO: 30) of bpTK7.

Detailed Description of the Invention

Novel protein tyrosine kinase genes have been identified, their nucleic acid sequences determined, and
10 the amino acid sequences of the encoded proteins deduced. The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology to
15 with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. These genes have been shown to be present in both megakaryocytic and lymphocytic cells.

To facilitate the isolation and identification of
20 these novel pTKs, two sets of DNA probes were used, as described in the Exemplification. The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2) (Matthews, W. Cell 65: 1143 (1991; Wilks, A. F. Proc. Natl. Acad. Sci. USA
25 86:1603 (1989)). These sequences were used as primers in a polymerase chain reaction to amplify tyrosine kinase DNA segments. (Mullis, K. et al., Cold Spring Harbor Symp. Advan. Biol. 51:263 (1986)).

The second set consisted of two oligonucleotide
30 sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4)

-8-

designed to amplify the nucleic acid sequence which encodes the highly conserved regions of the catalytic domains of the c-kit family of protein tyrosine kinases. These sequences were used as primers in the polymerase chain reaction in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these PTK primers were identified, isolated and subsequently sequenced.

In particular, fourteen PTK genes exhibiting significant homology with the c-kit subgroup of protein tyrosine kinases have been identified. Two PTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor) were identified in several megakaryocytic cell lines, including CMK 11-5, DAMI, UT-7 and UT-7 grown in erythropoietin, but not in the erythroid cell lines HEL, PMA stimulated HEL cells, or K562. Five PTK genes, referred to as LpTKs, were identified in lymphocytic, as well as in megakaryocytic cells. One PTK gene, referred to as HpTK5 was identified in human hepatoma cells and six genes, referred to as bpTKs, were identified in human brain tissue.

SAL-S1 (SEQ ID NO:6 and 18) encoded by the nucleic acid sequence SEQ ID NOS:5 and 17, exhibits significant homology with the FLT/FLK family of pTKs. SAL-D4 (SEQ ID NO:8) encoded by SEQ ID NO:7, is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3 (SEQ ID NO:12) encoded by the SEQ ID NO:11, is related to the NGF receptor family of pTKs. The remaining LpTKs, LpTK2 (SEQ ID NO:10) encoded by SEQ ID NO:9; LpTK4 (SEQ ID NO:14) encoded by SEQ ID NO:13; LpTK13 (SEQ ID NO:16) encoded by SEQ ID NO:15 LpTK25 encoded by SEQ ID NO: 22, also exhibit

-9-

sequence homology with known protein tyrosine kinases (Data not shown).

HpTK5 (SEQ ID NO: 24) encoded by SEQ ID NO: 23 and the bpTKs 1, 2, 3, 4, 5 and 7 (SEQ ID NOS: 25-30) respectively, also exhibit sequence homology with known protein tyrosine kinases.

Thus, as described above, DNA which hybridize with DNA encoding amino acid sequences present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases have been isolated and sequenced. These isolated DNA sequences, collectively referred to as pTKs genes, (and their deduced amino acid sequences) have been shown to exhibit significant sequence homology with known members of receptor tyrosine kinase families.

Once isolated, these DNA fragments can be amplified using known standard techniques such as PCR. These amplified fragments can then be cloned into appropriate cloning vectors and their DNA sequences determined.

These DNA sequences can be excised from the cloning vectors, labeled with a radiolabeled nucleotide such as ^{32}P and used to screen appropriate cDNA libraries to obtain the full-length cDNA clone.

The pTk genes as described above have been isolated from the source in which they occur naturally, i.e. megakaryocyte and lymphocytic cells. The present invention is intended to include pTk genes produced using genetic engineering techniques, such as recombinant technology, as well as pTk genes that are synthesized chemically.

The deduced amino acid sequences of the pTK genes include amino acid sequences which encode peptides exhibiting significant homology with the catalytic domain

-10-

of protein tyrosine kinases of the c-kit subgroup of tyrosine kinases. These proteins, encoded by the pTk genes, can include sequences in which functionally equivalent amino acid residues are substituted for
5 residues within the sequence, resulting in a silent change, that is a change not detected phenotypically. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent,
10 resulting in a silent substitution.

In addition, the protein structure can be modified by deletions, additions, inversion, insertions or substitutions of one or more amino acid residues in the sequence which do not substantially detract from the
15 desired functional tyrosine kinases properties of the peptide.

Modified pTKs of the present invention, with receptor tyrosine kinase activity can be made using recombinant DNA techniques, such as excising it from a vector containing a
20 cDNA encoding such a protein, or by synthesizing DNA encoding the desired protein mechanically and/or chemically using known techniques.

An alternate approach to producing the pTKs of the present invention is to use peptide synthesis to make a
25 peptide or polypeptide having the amino acid sequence of such a protein. The peptides or modified equivalents thereof, can be synthesized directly by standard solid or liquid phase chemistries for peptide synthesis.

Preferably, the pTKs of the present invention will be
30 produced by inserting DNA encoding the proteins into an appropriate vector/host system where it will be expressed.

-11-

The DNA sequences can be obtained from sources in which they occur naturally, can be chemically synthesized or can be produced using standard recombinant technology.

This invention also pertains to an expression vector
5 comprising a pTK gene of the present invention, encoding for a protein which exhibits receptor tyrosine kinase activity.

The pTK genes of the present invention can be used for a number of diagnostic and therapeutic purposes. For
10 example, the nucleic acid sequences of the pTK genes can be used as probes to identify other protein tyrosine kinases present in other cell types, including eukaryotic and prokaryotic cell types.

The nucleic acid sequences can be used to design
15 drugs that directly inhibit the kinase activity of protein tyrosine kinases, or to design peptides that bind to the catalytic domain of tyrosine kinases, thus inhibiting their activity. These sequences can also be used to design anti-sense nucleotides that can also inhibit, or
20 destroy, tyrosine kinase activity. Such inhibition of tyrosine kinase activity would be desirable in pathological states where decreased cellular proliferation would be beneficial, such as leukemias or other malignancies.

25 The nucleic acid sequences can also be used to design drugs, peptides or anti-sense nucleotides as above, but with enhancing, rather than inhibitory effects, on tyrosine kinases. Such enhanced tyrosine kinase activity would result in increasing the phosphorylation of
30 substrates (exogenous, as well as endogenous tyrosine residues). Enhanced effects would be desirable in states

-12-

where increased cellular proliferation would be beneficial, such as anemias, bleeding disorders and during surgical procedures.

The pTK genes of the present invention can also be
5 used to obtain soluble fragments of receptor tyrosine kinases, capable of binding their respective ligands (i.e. fibroblast growth factor).

PTK genes encoding soluble receptor tyrosine kinase fragments can be produced using recombinant DNA techniques
10 or synthetically. In either case, the DNA obtained encodes a soluble pTK fragment which lacks a substantial portion of the hydrophobic transmembrane region to permit solubilization of the fragment.

These soluble pTK protein fragments can be introduced
15 exogenously to act as competitors with the endogenous, membrane bound pTK for their respective ligands, thus inhibiting tyrosine kinase activity. Alternately, a modified soluble pTK protein fragment can be introduced which binds the ligand but does not activate kinase
20 activity.

These soluble pTK protein fragments can also be used in binding assays to detect ligands such as growth and differentiation factors. Once these ligands are identified, they may be altered or modified to inhibit or
25 enhance kinase activity. For example, the ligands may be modified or attached to substances that are toxic to the cell, such a ricin, thus destroying the target cell. The substance may be a super-activating substance which, after binding to the pTK, may substantially increase the kinase
30 activity, or activate other growth factors.

-13-

pTk genes of the present invention would also be useful to develop diagnostic tools for in vitro screening assays for ligands such as growth factors or differentiation factors that inhibit or enhance kinase activity. The proteins encoded by the pTK genes can also be used in such assays, or as immunogens to produce monoclonal or polyclonal antibodies to be used in such assays.

Such antibodies can also be used in methods of treating conditions in which an individual would benefit therapeutically if protein tyrosine kinase activity could be modified, such as increasing platelet production in bleeding disorders.

The present invention will now be illustrated by the following Exemplification, which is not intended to be limiting in any way.

Exemplification: The Identification and Isolation of the pTK Genes

To facilitate the isolation and identification of these novel pTK genes, two sets of DNA probes were used. (See Table).

The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2). These sequences were used as polymerase chain reaction (PCR) primers, using standard PCR techniques, to amplify tyrosine kinase DNA segments.

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4) selected from the highly conserved regions of the catalytic domains of the c-kit subgroup of protein

-14-

tyrosine kinases. These sequences were also used as polymerase chain reaction primers in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK
5 primers were identified, isolated and subsequently sequenced using known laboratory techniques.

TABLEFirst Round of Amplification

PTK1

10 CGGATCCACAGNGACCT

PTK2

GGAATTCCAAAGGACCAGACGTC

Second Round of Amplification

PTK3 (kit family specific)

15 CGGATCCATCCACAGAGATGT

PTKKW (kit family specific)

GGAATTCCTTCAGGAGCCATCCACTT

Equivalents

Those skilled in the art will recognize, or be able
20 to ascertain using no more than routine experimentation,
many equivalents to the specific embodiments of the
invention described herein. Such equivalents are intended
to be encompassed by the following claims.

-15-

CLAIMS

The invention claimed is:

1. Isolated DNA of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA
5 encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences
10 consisting of:
 - ✓a) SAL-S1 (SEQ ID NOS:5 and 7);
 - ✓b) SAL-D4 (SEQ ID NO:7);
 - ✓c) LpTK 2 (SEQ ID NOS:9 and 19);
 - ✓d) LpTK 3 (SEQ ID NO:11);
 - 15 ✓e) LpTK 4 (SEQ ID NOS:13 and 21); → MLK3
 - ✓f) LpTK 13 (SEQ ID NO:15);
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
3. Isolated DNA of Claim 1 which encodes an amino acid
20 sequence selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NOS:6 and 18);
 - b) SAL-D4 (SEQ ID NO:8);
 - c) LpTK 2 (SEQ ID NOS:10 and 20);
 - d) LpTK 3 (SEQ ID NO:12);
 - 25 e) LpTK 4 (SEQ ID NO:14);
 - f) LpTK 13 (SEQ ID NO:16);
 - g) HpTK 5 (SEQ ID NO:24);

-16-

- h) bpTK 1 (SEQ ID NO:25);
- i) bpTK 2 (SEQ ID NO:26);
- j) bpTK 3 (SEQ ID NO:27);
- k) bpTK 4 (SEQ ID NO:28);
- 5 l) bpTK 5 (SEQ ID NO:29); and
- m) bpTK 7 (SEQ ID NO:30).

4. Isolated DNA of human megakaryocytic origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.

10

5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:

- a) SAL-S1 (SEQ ID NO:6);
- 15 b) SAL-D4 (SEQ ID NO:8);
- c) LpTK 2 (SEQ ID NO:10);
- d) LpTK 3 (SEQ ID NO:12);
- e) LpTK 4 (SEQ ID NO:14); and
- f) LpTK 13 (SEQ ID NO:16).
- 20 g) HpTK 5 (SEQ ID NO:24);
- h) bpTK 1 (SEQ ID NO:25);
- i) bpTK 2 (SEQ ID NO:26);
- j) bpTK 3 (SEQ ID NO:27);
- k) bpTK 4 (SEQ ID NO:28);
- 25 l) bpTK 5 (SEQ ID NO:29); and
- m) bpTK 7 (SEQ ID NO:30).

-17-

6. A homogeneous protein of human megakaryocytic origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.
- 5 7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:
- ✓a) SAL-S1 (SEQ ID NO:6);
 - ✓b) SAL-D4 (SEQ ID NO:8);
 - 10 ✓c) LpTK 2 (SEQ ID NO:10);
 - ✓d) LpTK 3 (SEQ ID NO:12);
 - ✓e) LpTK 4 (SEQ ID NO:14); and
 - ✓f) LpTK 13 (SEQ ID NO:16).
 - ✓g) HpTK 5 (SEQ ID NO:24); *HTK, MDK2, RET*
 - 15 ✓h) bpTK 1 (SEQ ID NO:25);
 - ✓i) bpTK 2 (SEQ ID NO:26);
 - ✓j) bpTK 3 (SEQ ID NO:27);
 - ✓k) bpTK 4 (SEQ ID NO:28);
 - ✓l) bpTK 5 (SEQ ID NO:29); and
 - 20 ✓m) bpTK 7 (SEQ ID NO:30).
 - ✓ *LpTK 25, FLT3*
8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.
9. A protein of Claim 8 encoded by the nucleotide
25 sequence (SEQ ID NO:5).
10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

-18-

11. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:17).
12. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO: 18).
- 5 13. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FGF receptor family of protein tyrosine kinases.
14. A protein of Claim ¹³10 encoded by the nucleotide sequence (SEQ ID NO:7).
- 10 15. A protein of Claim ¹⁰10 encoded by the amino acid sequence (SEQ ID NO:8).
16. A protein of human megakaryocytic origin which exhibits significant sequence homology with the NGF receptor family of protein tyrosine kinases.
- 15 17. A protein of Claim ¹⁴14 encoded by the nucleotide sequence (SEQ ID NO:11).
18. A protein of Claim ¹⁴14 encoded by the amino acid sequence (SEQ ID NO:12).

-19-

19. A DNA expression vector containing a DNA sequence of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.
- 5
20. The DNA expression vector of Claim 17^{19?} containing a DNA sequence selected from the group consisting of:
- 10
- a) SAL-S1 (SEQ ID NO:5);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NO:9);
 - d) LpTK 3 (SEQ ID NO:11);
 - e) LpTK 4 (SEQ ID NO:13); and
 - f) LpTK 13 (SEQ ID NO:15).
- 15
- g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
21. A cell transformed by the expression vector of Claim 17.

1/35

sa/s1 (160 bases)
FLKI-LIKE

1	PTKI/3 PRIMERS	21	41
5' <u>ggatcctgtgcatcagtgacttagggctagggaacattctgctgtcggaaagcga</u> <u>ctggt</u>			
		D P V H Q a L R A R N I L L S E S D V V	
61		81	101
gaagatctgtgactttggccttggccctggccgggacatctacaagagaccccgctacgtccgcaa			
		K I C D F G L A R D I Y K D P S (Y) V R K	
121		141	PTKKW PRIMER
gcagcccggtgccctgaagtggatggcgccagaattc 3'			
		H A R L P L K W M A P E F	

FIGURE 1

sald4 (147 bases)-
FGFR-LIKE

1	PTK1/3 PRIMERS	21	41
6'	<u>ggatccattcacagagaccctagcagcacgcaacatcctgggtctcagaggaccctggtacc</u>		
	G S I N R D L A A R N I L V S E D L V T		
61		81	101
	<u>aaggtcagcgactttggccctggccaaagccgagcggaagggttagactcaagccggctg</u>		
	K V S D F G L A K A E R K G L D S S R L		
121	PTKKW PRIMER	141	
	<u>ccgtcaaatggatggctcccgattc</u> 3'		
	P V K W H A P E F		

FIGURE 2

3/35

LpTK2

GTTGGAATTCCTTCCGGCGCCATCCATTTACCGGCAGCTTTATTTTCGTGTCTAGATTCA
TAGATGTCTTCATTATCTACCTTAAAACTCTGGCAAGTCCAAAATCTGCTACTTTGTAG
ATATTATGTTACCAACGAGGACATTCCT

FIGURE 3A

LpTK3

GTGCACAGGGATCTCGCGGCTCGGAACATCCTCGTCGGGGAAAACACCCCTCTCGAAAGTT
GGGGACTTCCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCATGACCACAAT
ATCCCCTACAAATGGATGGCCCCCTGAGGGAA

FIGURE 3B

LpTK4

GTTACCGAGATCTCAAGTCCAACAACATTTTGCTGCTGCAGCCCATGAGAGTGACGAC
ATGGAGCACAAGACCCTGAAGATCACCGACTTTGGCCTGGCCCGAGAGTGGCACAAAACC
ACACAAATGAGTGCCGC

FIGURE 3C

LpTK13

GTCAATCGTGACCTCGCCGCCCCGAAATGTGTTGCTAGTTACCCAACATTACGCCAAGATC
AGTGATTTCCGACTTTCCAAAGCACTGCGTGCTGATGAAACTACTACAAGGCCAGACC
CATGGAAAGTGGCCTGTCAAGTGGTACGCTCCGGAATGCATCAACTACTACAAGTTCTCC
AGCAAAAGCGATGTCTGGTCCTTTGGAATTC

FIGURE 3D

1 TTCGAGCTCG CCGGACATTG ATTATTGACT AGTTATTANT AGTAATCAAT TACGGGGTCA TTAGTTCAAT GCCCATATAT GGAGTTCCGC GTTACATATAC
AAGCTCGAGC GGGCTGTAACTAATATACTCA TCANTATTA TCATTAGTAA ATGCCCCAGT ATCAAGATAT CCGGTATATA CCTCAGGCG CANTGTATTG

101 TTACGTAAT TGGCCCGCT GCGTACCGC CCACGACCC CCGCCCATTT ACCTCATTA TCAGCTATGT TCCCATAGTA ACCCAATAG GGACTTTCCA
AATGCCATT ACCGCGCGA CCGACTGGC GGTTCCTGG GGGGGTAAC TCCAGTATTT ACTGCATACA AGGTATCAT TCGGTATATC CCTGAAGGT

201 TTGACCTCAA TGGGTGGAGT ATTACGGTA AACTGCCCAC TTGGCAGTAC ATCAAGCTA TCATATGCCA AGTACCCCC CTATTCACGT CAATGACGGT
AACTGCAGT ACCCAGCTCA TAAATGCCAT TTGACGGGTG AACCTCATG TACTATACGT ACTATACGGT TCATGCGGGG GATAACTGCA GTTACTGCCA

301 AANTGGCGC CCTGGCTTA TCCCGATAC ATGACCTAT GGGACTTTC TACTTGGCAG TACATCTACG TATTAGTAT CCGTATTACC ATGGTATGC
TTTACCGGC GACCCGTANT ACGGTGATG TACTGGAATA CCTGGAAGG ATGAACGCTC ATGTAGTGC ATATATGTA CCGATATAG TACCCTACG

401 GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTCGA CTCACCGGGA TTTCATAGTC TCCACCCCAT TCACGTCAAT GGGAGTTGT TTTGGCACA
CCAAACCGT CATGTAGTA CCGGCACCTA TCGCCAAACT GAGTGCCTT AAGGTTTCTG ACTGCATTA CCTCAACA AACCTGTCT

501 AATCAACGG GACTTTCCA AATGTCTTA CAATCCGCC CCAATGAGC AATGGCGG TAGGCTGTA CCGTGGGAGG TCTATATAG CAGAGCTCT
TTAGTTGCC CTGAACGTT TTACAGCAT GTTGAGGCG GTTACTCTG TTTACCGGC ATCGGCAT GGCACCTTC AGTATATTC GTCTCGACA

601 TTAGTGAAC CTCAGATCC CTGGAGAGC CATCCACCT GTTTGACCT CCATAGACA CACCGGACC GATCCAGCT CCGCGGCGG GAACGTGCA
AATCACTTG CAGCTATGG GACCTCTGG GTAGTGGA CAAACTGGA GGTATCTCT GTGGCCCTG CTAGGTCCA GCGCGCGCG CTTCCTACCT

FIGURE 4A

5/35

701 TTGGACGGC GATTCUUGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA GTCTATAGGC CCALTGGCT TCCTTAGNAC CCGGTACAA TTAAACAA
AACCTGGC CTAAAGGGCA CGGTTCTAC TGCATTATG CCGGATATCT CAGATATCG GTGAAACCA AGCAATCTG CCGCATGTI AATTATGTAT

801 ACCTATGTA TCATACACAT ACCATTAGG TGACACTATA GAATAACATC CACTTGGCT TTCTCTCCAC AGGTGTCAC TCCAGGTCC AACTGCACCT
TGGATACAT AGTATGTGA TGTAAATCC ACTGTGATAT CTATTGTAG GTGAAACCA AAGAGAGTG TCCACAGGTG AGGTCCAGG TTGACGTGA

901 CGGTTCTATC GATTAATTC CCGGGGATC CTCAGAGAT CCTCGACCT CGAGATCAT TGTGCTGGC CGATTCTT ATCACTGATA AGTTGGTGA
GCCAGATAG CTAACCTAAG GGGCCCTAG GAGATCTTA GAGATCTTA GAGATCTTA ACACGACCG GCTTAGGTA TACTGACTAT TCAACCACT

1001 CATATTATGT TTATCAGTA TAAAGTGTA AGCATGACAA AGTTCAGCC GAATACAGTG ATCGTGGC CCTAGACCT GTTGACGAG GTGGCGTAG
CTATAATACA AATAGTCACT ATTTCACAGT TCGTACTGT TCAACGTGG CTATGTCAC TAGGCACGG CGCATCTGA CAATCTGTC CAGCGCATC

1101 ACGGTCTGAC GACAGCGAA CTGGCGGAC GTTGGGGT TCAGCAGCG CGCTTTACT GGCATCTCA GAACNAGCG CGCTGCTCG ACGCACTGG
TCCACAGCTG CTGTGCGTT GACCGCCTG CCAACCCCA AGTCTGCGC CGCGAATGA CCGTGAAGTC CTGTGCGC CGCAGCAGC TCGGTGACG

1201 CGAAGCCATG CTGGCGGACA ATCATAGCAC TTGGTGGCG AGAGCCGACG ACGACTGGG CTGATTCTG ACTGGGATG CCGCAGCTT CAGGCAGCG
GCTTGGGTAC GACCGCTCT TAGTATCGT AGCCACCGG TCTGGCTCC TCGTACCGC GAGTAAGAC TGACCTTAC GCGCTGGA GTCCGTCCG

1301 CTGCTCGCT ACCGCCAGCA CAATGGATCT CGAGGGATCT TCCATACCTA CCAGTTCTG GCTGAGGT CGGGGCGCA CTACTCTTG ATGTATTACT
GACGAGCGA TGGCGGTCT GTTACCTAGA CTCCTCTAGA AGGTATGAT GGTCAAGACG CGGACGTCA GCGCCGCGT GATGAGAAC TACATATGA

FIGURE 4B

1401 CATATTACCA AGCAATAACT GCGGGGCACA GGTTCAGGTG CTGAAGGGAC ATTGTGAGAA GTGACCTAGA AGGCAAGAGG TGAGCCCTCT CTCACGCTGG
 GTATAATGGT TCCTTATTGA CCGCCCGTGT CCCAGTCCAC GACTTCCTC TAACTACTTT CACTGGATCT TCGTTCTCC ACTCGGGAGA CAGTGGGACC

1501 CATAGGGCC GCTTGAGGCG TCTTTGTGCA AGCAGTAACG CCAGTGTCTG GUANGGCACC TGTACTCAG CAGACCATCA AAGGCGCTCT CCCTTTCTTT
 GTATTCGCG GCACTCCCG AGNANCCAGT TCGTCATTCG GGTACACAGC CTTCCCTGG ACAATGACTC GTCTGTACTT TCCCGCAGA GGAAGAAGAA

1601 GCAGCAGTCA GGGAACTCT TCTTCCACCA GCTTCTTCTG GGAGGTGGA TATTATCAG CCTGCCCG AGTCATCCCG AGGCTAACCC CTCCTCTCTG
 CTTCTCACT CCTTGTGAG ACGAGTGTCT CGAAGAACAC CTCGCACT ATANTAGCTC CGGACGGCGG TCAGTNGCC TCCGATGG GGAGGGACAC

1701 GTGCTTCACT GGTACACTC CTGTTCACCT TCTATGCTCC TCTTGCTCTC CTGTTCTCTC TTGGAAGTTT GTAGTAGATA GCAGAAGAA TAGGCAAGT
 CACGAAGTCA CCAGTGTGAG GACAGGTGA NGTAGCAGG AGAACGGG GACCAAGG ACCTTCANA CATCACTAT CGTCTCTTT ATCCCTTCA

1801 CTTAAGTCT TTGATCTTC TTATAGTCC AGAGAAGAA TGCTGACGTA TGCTGCTTC TCTCTCTG CTTGAGTAC CTGAAGCGC TTCTTGTCT
 GAATTCGGA AACTAGNAG ATATTTCAGG TCTCTCTTT AGACTGCAAT AGCAGGAG AGAGAGAGC GAAGTGTG GACTTGGCG AAGAACAGA
 349 O R P G S E O R

1901 ATACCTGCTC TCTATCTGCT CACACTCTCT CGAGGCGAGC ACCATCCAC TCTCTCTCTG GTTGTCCACA GAGCTTTGT AGTCTGTCG GGTCTGTCG
 TATGAGGAG AGATAGACCA GTCTGAGGAG CTCTGCTCTG TGGTAGGCTG ACAGACAGC CAACAGGTCT CTCGGAACA TCCAGAACCC CCACTACTCC
 341 Y R S E I Q E C E E S A L V N G S D Z Q N D V S G K Y Z T P Z M P

2001 AATTCCTCAA ATGTCTTCAT CCTGAGGAA CCACGGTCT CAGCCCTCT GGCAGGCAC CCGGAAGG ACACCAAGT GTATACCTG GCGCCAGCC
 TTAAGGAGT TACAGAGTA GCACTCTCT GTGCCGAGA CCGTCCGTG GGCCTTCTC TGTGGTCAA CATATGGAC CGCGGTCCG
 308 F E E F T K M R S S G R T E A G R A L C G P F S V W N Y Y R A A L

FIGURE 4C

2101 TGTGGCGCTG CAGGCTTGGC GGGCTGTCTT CAGGCTCAGC GTGGGCGATG TGTAGGGCCA TGTGGGACAC GTGGGAGAG GTCCCTCTT CTGAGCTCTG
 ACACCGCGAC GTCCGACCG CCGACAGTCG GACCGCTAC ACATCCGGT ACACCTGTG GACCTCTTC CACGGAGAA GACTCGAGAC
 215 S H R O L S P P S D E A D A Q A I H L A M T S V Q S F S G E E S S O

2201 AGAGCTGGC GGGGCAATGC AGACTCTCTC TTCTCTTGC AGCCCTCTC CCTGGAGCAG GTCCCCGAGG ATCTCCACCA GTCCGAGAA TGCAGGTCTC
 TCICGACCGG CCCCCTTACG TCTGGAGAG AGGAGACAG TCCGGGACG CAGCTCTG G O L L D G L I E V L E S F A P R
 241 S S R P A M C V E E E E O L G R G O L L D G L I E V L E S F A P R

2301 GCCTTGGGT CTCGACCA GCATTCAGC ATGATGCGG GTATGGCGG AGTGGCGCC TCCGGGCGG TCATCTTGT CCGTCTCTC AGCGCTGCG
 CCGAACCCCA GAGGCTGGT CGTCAAGTCG TACTAGCCG CATACGCCG TCACCGGTG AGGCCCGG AGTAGGACA CCGCAGAG TCGGCGACG
 208 A R P D G S W C N L M I R R I A P T A L E P A R H R T G D R L R Q

2401 AGAATCTCTC ATTGATCTGC ACCGAGGGT ACGGGAGCG CCCAGAGAG AGATCTCC ACAGAGCAG CCCAAGGAC CACAGTCAC TCTGGTGGT
 TCTGAGAG TACTAGACG TGGGTCTCA TCCCTCTCG GGGTCTCTC TCTAGAGG P I E W L L V G F S W V D S Q T T
 175 C F E E N I O V G P Y P S A G L S P I E W L L V G F S W V D S Q T T

2501 GTACALTTG TCGAGATGC TTTCAGGCG CATCCACTC AGGGCAGCG GGCACCTGCC CTTCGGGAG TACTCGGGT CTTGTAGAT GTCCCGACA
 CATGTGGAC AGCTCTACG AAGTCCCG GTAGGTGAG TCCCTCTCG CCGTGACG GAGGCTGC ATCAGCCCA GAACATCTA CAGGCGCT
 141 Y V R D F Y S E P A H W K L P L R A S G K R V Y D P D K Y I D R A

2601 AGGCCAAGT CACAGATCTT CACACCTCG CTTTCGACA GCAGATGTT CCGAGCAGC AGGTCTCTT GATGCACTT TCGGAGCC AGGAATCCA
 TCCGTTTCA GTGTAGAA GTGTGACG GAAGGCTGT CGTCTTACA GGTCTCTCG TCGAGACA L D R M I C K N S A L P E
 108 L C F D C I K V V D S E S L L I M R A A L D R M I C K N S A L P E

2701 TCCCTTGGC CACCTGGAG CTGTAGCAGA CAAGATCTC CATGCTCAGC GGGTCAGC ACAGTCTCT AGCTTCTTG TCTGGAGAG CCGGCTCTCG
 AGGGAGCG GTGGACTTC GACATCTCT GTTCTAGAG GTACCTCTG CCGAGCTCG TGTCCAGAG TCGAGAAC AGACTCTCT GGGCGAGCG
 75 M G R A V O P S Y C V L D E H T L P S L W L P E A E Q D P S A R R A

FIGURE 4D

8/35

FIGURE 4E

2801 TCCGCCCTCG GTCTTCAGAA ACCGCGCAAA GAGGACCTG TCGTGTCTCC CCGGCCGCT CCGATCCAGC CTGGCGAGCT CCACCATGC GCGGAGCGGT
 AGCGGGAGC CAGAGCTCT TGGCGCTT CTCCTGGAC AGCGAGGAG GCGCGCGGA GGTAGTGG GACCGCTGA GCGGTACG CCGCTTCGA CCGCTTCGA
 41 C G E T K S P R A F L V R D S S G P R R R D L R A L E V N A R P R
 2901 CCGCGCTGCT CCGGAGACTT CTCCTGCGGA TGCACGAAC TCGCTCGAG GCGCCAGTC GTCCGCGCA GAGCGCTC CATTCCTCG CATTCCTCG CCGCGCGCG
 GCGCGGACGA GCGCTCTGA GAGGAGCTT AGTGTCTTC ACCGAGCTCC CCGCGCTCG CAGCGCGCT CTCCGCGAG GTAGCGGCG GCGCGCGCG
 6 G R Q E P S K E
 3001 CCGCCCGCAG CCGCGCGCT CACCGGCGAG GGGCTCGGC CCGACTCTA GAGTCGACT GCAGAGCTT GGCCTCATG GCGCCATG TTTATTCGAG
 GCGCGGCTC CCGCGCGCA GTGGCTGCT CCGAGCGGC GCGCTCAGT CTCAGCTGA CCGCTTCGA CCGCGGTAC CCGGTTCAG NANTACGTC
 3101 CTTATATGG TTACAAATA AGCAATAGCA TCACAAATT CACAAATAA GCATTTTT CACTGCTTC TAGTGTGGT TTGTCCAAAG TCATCAATGT
 GAATATTACC AATGTTTTC TGTATTCGT AGTGTATAA GTGTTTTC CTAATAAATA GTAGCTTAG ATCAACACCA AACAGTTTG AGTAGTTACA
 3201 ATCTTATCAT GTCTGATCG ATCGGAAAT ATTCGGCGC AGCACCATGG CCGCAATAA CCGTGAAG AGCACTTC TTAGCTACTT TCTGAGCGCG
 TAGAATAGTA CAGACTAGC TAGCCCTTA TTAAGCGCG TCGTGTACC GCACTTAT GCGACTTC TCGTGAAC ATCCATGA AGACTCCGC
 3301 AAAGAACCAG CTGTGGAATG TGTGTCACTT AGGTGTGA AATTCGCGAG CCGCGCGAG AGCGAGAT ATGCAAGCA TGCATCTCA TTAGTCAGCA
 TTTCTTCTC GACACCTTAC ACACAGTCA TCCUACACT TTCAGGGGTC CGAGGGGTC TCGTGTTC TAGGTTCGT AGTAGAGT ANTACGTCT
 3401 ACCAGGTGT GAAAGTCGC AGCTCCCA GAGGCGA GATGCAAG CATGCTTC AATTAGTCA CAACATAGT CCGCGCGTA ACTCCGCGCA
 TGTCCACAC CTTTCAGCG TCCGAGGCT CCGCTCTT CATACGTT CAGCTAGT TATTCAGT GTTGATCA GCGCGGGAT TCAGCGCGGT

9/35

FIGURE 4F

3501 TCCGCGCCT AACTCCGCC AGTCCGCC ATTCTCGCC CATGGCTGA CTAAATTTT TATTTATGC AGAGGCCGAG CGCGCTCGG CCTCTGAGCT
AGGGCGGGA TTGAGGGGG TCAAGGGGG TAAAGGCGG TAAAGGCGG GUTACCGACT GATTAAANA ANTAANTAGC TCTCGGCTC CGCGGAGCC GGAGACTCGA

3601 ATTCCAGAG TAGTGAGGAG GCTTTTGG AGGCTAGGC TTTCGAAA AGCTGTAG AGCTGGCAC TGGCGTGTG TTACAAAGT CTTGACTGGG
TAAAGTCTC ATCACTCTC CCAANAAAC TCCGNTCC ANAACCTTT TCGAANTG TCGAACCTG ACCGGACCA AATGTTCGA CCACTGACCC

3701 AAACCCCTG CGTACCCAA CTAAATGCC TTCCAGACA TCCCCCTTC GCCACTGGC GTAAAGCA AGAGGCCCG ACCGATCCC CTTCCCAAACA
TTTGGGACC GCANTGGT GANTTAGCG AACGTCTGT AGGGGGAAG CGGTGACCG CATTAATCT TCTCGGCGG TGGTACCGG GAAGGTTGT

3801 GTTCGGTAG CTGANTGGG ATGGCGCT GATCGGTAT TTCTCTTA GGCATCTGT CGGTATTC CACCGCTAC GTCAAGCA CCAATAGTACG
CAACGCATC GACTTACCG TTACCGGA CTAGGCATA AAGAGGAT GGTAGNAC CGGTAAAGT GTGGGTATG CAGTTCTG TGTATCATC

3901 CGCCCTGTAG CGCGCATTA AGCGGCGG GTGTGTGTGT TACGCGCAG GTCAACGCTA CACTTGGCA GCGCTAGCG CCGCTCTCT TCGCTTCTT
CGCGGACATC GCGCGTAT TCGGCGCC CACACGACA ATGCGGTG CACTGGCAT GTCAACGCT CCGGATCG CGCGAGGA AGCGAAGA

4001 CCTTCTTT CTGCGCACGT TCGCGCGCT TCCCCGTCA GCTTAATC GGGGCTCC TTGAGGTC GATTTAGTG CTTACGCA CCTCGACCC
CGAAGCAA GACCGTGCA AGCGCGCA AGGGCAGT CGAGTTAG CCGCGAGG AATCCNAG CTAATCAC GAATCCCT CGAGCTCGG

4101 AAANAATTC ATTGGCTCA TGGTACAGT AGTGGGCA CCGCTGATA GACGTTTT CCGCTTCA CGTGGAGTC CAGTTCTT AATAGTGAC
TTTTGAC TAAACCACT ACCAAGTCA TCACCGGT TCGGACTAT CCGGCAAA CCGGCAACT GCACCTCAG GTCAAGAA TTATCACCTC

10/35

FIGURE 4G

4261 TCTTGTTCCA AACTGGAACA AACTCAACC CTATCTGGG CTATCTTTT GATTATAAG GATTTGGCC TATTGGTAA AAATGAGCT
ACAACAGGT TTGACCTGT TGTGACTGG GATAGAGCC GATAGAAA CTAATATTC CTAAGCCG ATAACCAAT TTTTACTCGA

4301 GATTTAACAA AAATTAACG CGAATTTAA CAAATATTA ACCTTACAA TTTATGGG CACTCTAGT ACAATCTGT CTGATGCCG ATACTTAAGC
CTAATGT TTTAANTGC GCTAANAAT GTTTATAT TGCANAATG AATAACCA CTAAGAGTCA TATTAGACGA GACTAGCCG TATCAATTCG

4401 CAATCCCT ATCGTAGCT GACTGGTCA TGGCTGGCC CGACACCCG CCACACCCG CTGACCGCT TGTCTGCTCC CGGATCCCG
GTTGAGGCA TAGCATGCA CTGACCCAGT ACGGACCGG GCTGTGGC GTTGTGGC GACTGCGG GACTGCGG ACAGACGAG GCGTAGCGG

4501 TTACAGACAA GGTGTGACG TCTCCGGAG CTGCATGT CTGAGGTTT CAUCGTATC ACCGAACG CCGAGGCACT ATCTTGAG AGGAAGGCG
AATCTCTG CGACACTGC AGAGGCCCTC GAGGTACCA GTCTCCAAA GTGGCAGTAG TGGCTTGG CCTCCGTC TAAGNACTT TCGTTTCCC

4601 CTCGTGATC GCTATTTT ATAGCTAAT GTCATGATAA TAATGTTTC TTACAGTCA GGTGGACTT TTGGGGAAA TGTGGCGGA ACCCTATTT
GAGCACTATG CCGATAAAA TATCCAAATTA CAGTACTAT ATTACCAAG ATCTGCAGT CCACCGTCA AGCCCTTT ACACCGCTT TGGCATAA

4701 GTTATTTT CTAAATACAT TCAATATGT ATCCGCTCAT GAGCAATAA CCTGATAA TGCTCAATA ATATTGAAA AGGAAGTAT TCAGTATTC
CAATAAAA GATTATGTA AGTTATACA TAGCGAGTA CTCTGTATZ GCGACTATT ACGAAGTAT TATNACTTT TCTTCTCAT ACTCATAGT

4801 ACATTTCCGT GTCCGCTTA TTCCCTTTT TCGCGCATTT TCGCTTCTG TTTTGCTCA CCGAGAACG CTGTGAAG TAAAGATCC TGAAGATCAG
TGTAAAGCA CAGCGGAT NAGGNAAAA ACGCGTAA TCGCGCATTT TCGCTTCTG GAGCACTTC ATTCTTACG ACTTCTAGT

11/35

FIGURE 4H

4901 TTGGGTGCAC GAGTGGGTTA CATCGAACTG GATCTCAACA GCGGTAAAGT CCTTGAGAGT TTTCGCCCCG AAGAACGTTT TCCAAATGATG AGCACTTTT
AACCACCGTG CTCACCCAAAT GTAGCTTGAC CTAGAGTTGT CCGCACTCTA GCAACTCTCA AAGCGGGGC TTCTTCNA AGGTACTAC TCGTGAAAA

5001 AAGTCTGCT ATGTGGCCG GTATTATCC GTGATGACG GGGCAAGAG CAACTGGTC GCGGCATACA CTATTCTCAG AATGACTTG AATGACTTG
TTCAAGNCGA TACACCGGC CATATAGGG CACTACTGG GCCGCTCTC GTTGACCCAG CCGGTATGT CATANGTC TTAAGTACC AACTCATGAG

5101 ACCAGTCACA GAAAGCATC TTACGGATG CATGACAGTA AGAATATAT GCAATGCTG CATACCATG AGTATACA CTGCGCCAA CTACTTCTG
TGTGAGTG CTTCGTAG AATGCTACC GTACTGTCT TCTTATTA GGTACAGAG GTATTGTG TCACTATGT GACCGCGGT GATGAGAG

5201 ACAACGATCG GAGGACGCA GGAGCTAAC GCTTTTTC ACACATCGG GGATCATGTA ACTGCGCTG ATCGTTGGG ACCGAGCTG AATGAAGCA
TGTTCGAC CTCTGCTT CTTCGATGG CGAAAAAC TGTGTATCC CCTATGAT TGAAGGAG TACCAACCT TGGCTGAC TACTTCTG

5301 TACCAACGA CGAGCGTAC ACCAGCATG CAGCAGCAAT GGCACAAAG TTGCGCAAG TATTAACTG GAACTACTT ACTTACTT CCGCGCAACA
ATGGTTGCT CTTCCACTG TGTGCTAG TGTGCTTA CCGTCTGCTT ACCTGCTT ATATGAC CTTGATGAA TGAGATCGA TGGCCGTTG

5401 ATTAATAGAC TGGATGAGG CCGATMAAT TGCAGGACCA CTTCGCTT CCGCTCTCC GCGGAGAG CCGACCGAC AATATACG TATTAGAC TCGCCACTG
TAATATCTG ACCTACCTG CCTATTCTA ACCTCTCTG GAGAGCGCA GCGGAGAG CCGACCGAC AATATACG TATTAGAC TCGCCACTG

5501 CGTGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAGCCCTC CCGTATCTA GTTATCTACA CGACGGGAG TCAGGCACT ATGATGAG
GCACCCAGAG CGCCATAGTA AGTCTGTAC CCGGTCTAC CATTCGGAG GGCATAGCAT CAATAGATG CTGCCCCC ACTCCGTTG TACTACTTG

FIGURE 41

5601 GAATAGACA GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGCTAA CTGTCAGACC AAGTTACTC ATATATACCT TAGATTGATT TAAACTCTT
CTTATCTGT CTAGCGACTC TATCCACGGA GTGACTAATT CGTAACCACT GACAGTCTGG TCAANTGAG TATATATGAA ATCTAACTAA ATTTTGAA

5701 TTTTAAATT AAAGGATCT AGGTGAGAT CCTTTTGTAT AATCTCATGA CCAAAATCC TTAACGTGAG TTTCTGTC ACTGAGGTC AGACCCGTA
AAATTAATA TTTTCTAGA TCCACTTCTA GGAANAATA TTAGAGTACT GGTTTTAGG AATTGCATC AAAGCNAAG TCACTGCGAG TCTGGGCAT

5801 GAAAGATCA AAGATCTTC TTGAGATCT TTTTCTGC GCGTAATCTG CTGCTGCAA ACATAAATC CACCGTACC ACCGTGGTT TGTTCGCGG
CTTTCTAGT TTCTAGAG AACTTAGGA AAAAAGACG CGCATTAGC GACGAAGCT TGTTTTTG GTGGCAGG TCGCCACCA ACNAACGGCC

5901 ATCAAGACT ACCAATCTT TTTCCGAGG TAATGCTT CAGCAGACG CAGATACCA ATACTGCTT TCTAGTGTAG CCGTAGTATG GCCACCACTT
TAGTCTCGA TGGTTGAGA AAGGCTTC ATTGACCGA GTGCTCTCC GTCTATGCT TATGACAGGA AGTCATC ATC GCATCATC CCGTGTGA

6001 CAAGAACTT GTAGCACCG CTACATACCT CGCTCTCTA ATCTGTATC CAGTGGCTG TCCAGTGGC GATAAGTCT GTCTTACCG GTTGACTCA
GTCTTGAGA CATCGTGGG GATGTATGA GCGAGACGAT TAGGACAAATG GTCACCGAG ACGGTCAACG CTATTACGA CAGANTGGCC CAACCTGACT

6101 AGACGATAGT TACCGGNTA GCGCGAGCG TCGGGCTGAA CCGGGGCTC GTGCACACG CCCAGCTTG ACCGACGAC CTACACCGAA CTGAGATACC
TCTGCTATCA ATGCGCTATT CCGGCTGCC AGCCGACTT GCGCCGACG CACGTGTGC GGTGGAAC TCGCTGCTG GATGTGCTT GACTCTATG

6201 TACAGCGTGA GCATTGAGA AGCGCCACG TTCCGGAAG GAGAAAGCG GACAGGTATC CCGTAAGCG CAGGTCGA ACAGGAGAG GCACGAGGG
ATGTCGCACT CGTAATCTT TCGCGGTGG AAGGCTTCC CTCTTCCG CTGTCCATAG GCAATTCGC GTCCAGCT TGTCTCTCG CGTCTCTCC

13/35

FIGURE 4J

6301 GCTTCCAGG GGAACGCT GGTATCTTA TAGTCTGTG GGGTTTCGC ACCTGCACT TGAGCTGCA TTTTGTGAT GTCGTCAGG GGGCGGGAGC
CGAAGGTCC CTTTGGCA CCAAGAAAT ATCAGGACAG CCCAAGCGG TGGAGACTGA ACTGCACTA ANAACACTA CGAGCAGTCC CCCCCTCTG

6401 CTATCGAAA ACCCAGCAA GCGGCGCTT TTACGTTTC TGGCTTTTG CTGGCTTTT GCTCACATGT TCTTCTCTG GTATGCTCT GTTCTGTCTG
GATACCTTT TCGGTCGT GCGCGGAAA ATGCGCAAG ACCGGAAC GACCGGAAA CGAGTGACA AGAAGGAGG CAATAGGGA CTAAAGACCC

6501 ATAACCTAT ZACCGCTTT GAGTGAGTG ATACCGCTCG CCGAGCGGA ACGACCGAGC GAGCGAGTC AGTGAGGAG GAAGCGAAG ACCGCCAAT
TATTGGCATA ATGGCGAAA CTCACTCGAC TATGGCGAG GCGGTGCTT TGTGCTCG GGTGCTCAG TCACTGCTC CTTCGCTTC TCGCGGTTA

6601 ACGCAACCG CCTCTCCCG GCGTTGGLC GATTCATTAA TCCAGCTGC ACGACAGGT TCCGACTGG AAGCGGGA GTGAGCGAA CGCAATTAT
TGGCTTTC GAGAGGCGC GGCACACCG CTAAGTAAT AGGTGACCG TGTGTCCAA AGGCTGACC TTTCGCTCGT CACTGCTT GCTTAATTA

6701 GTGAGTIACC TCACTCATTA GGCACCCAG GCTTACACT TTATGCTTC GCTCGTATG TTGTGCGAA TTGTAGCGG ATACCAATT CACACAGGA
CACTCAATG AGTGATAT CCGTGGGTC CGAATGTG ATACGATG CCGAGCATAC AATCACTT AATCTGCGG TATGTGTA GTGTGCTCT

6801 ACAGCTATGA CCATGATTAC GAATTAA
TGTGATACT GGTACTAATG CTAAAT

14/35

FIGURE 5A

1 TTGGAATCTCG CCGACATATG ATATATGACT AGTATATTAAT ACTAATACAT TACGGGTCA TTAGTTTATA GCCCATATAT GGACTTCCGC GTTACATATAC
 AATCTCAGC GGGCTGTAACT TAATAATCTA TCAATTAATTA ATGCCCCAAT ATCAAGTAT CCGGTATATA CCTCAGAGCG CATTGTATTG

101 TTACGGTAA TTGCGCGCCT GGCTGACCG CCAAGGACCC CCGCCCATTTG AGTCATATAA TGACGTATGT TCCCATAGTA ACCCAATAG GGACTTTCCA
 AATGCCATTT ACCCGGCGA CCGACTGCG GGTGCTTCTG GCGGGTAACT TGCAGTTATT ACTGCATACA AGGTATCAT TCGGTATATC CCTGMAAGGT

201 TTGACATCA TTGCTGGAGT ATTTACGGTA AACTGCCAC TGGCAGTAC ATCAAGTATA TCATATGCCA AGTACCCCC CTATTACGT CAATGACCGT
 AACTGCAGTT ACCCACCTCA TAAATGCCAT TTGACGGGTG AACCGTCATG TAGTTCACAT AGTATACGGT TCATGCGCGG GATNACTGCA GTTACTGCCA

301 AATGCGCCG CCTGGCATTA TGCCAGTAC ATGACCTTAT GGGACTTTC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGCTGATCG
 TTTACCGGC GGACCTAAT ACCGTCATG TACTGGAATA CCTGMAAGG ATGACCGTC ATATAGATGC ATATCAGTA GCGATNATCG TACCCTACG

401 GGTTTTGGCA GTACATCAAT GGGCTGGAT AGCGTTTGA CTCACGGGA TTCCAAATG TCCACCCCAT TGAGTCAAT GGGAGTTTGT TTTGGCACCA
 CCAAAACCGT CATGTAGTTA CCGCACCTA TCGCCAACT GAGTCCCTT AAGGTTTCA AGTGGGTA ACTGCAGTTA CCTCANACA AAACCGTGGT

501 AATCAACGG GACTTTCCAA ATGTCTGTA CACTCCGCC CCATTGACCG AATGCGCGG TAGGCTGTA CCGTGGAGG TCTATATAG CAGAGCTCGT
 TTTAGTTGCC CTGAAAGGTT TTACAGCATT GTTGAGCGCG GGTAACTGCG TTACCGCGC ATCCGCACAT GCCACCTCC AGATATATTG GTCTCGAGCA

601 TTAGTGACC GTCAGATCG CTGAGACCG CATTCCAGCT GTTTTACCT CCATAGAAGA CACCGGACC GATCCAGCTT CCGCGCCCG GAACGGTGA
 AATCACTTG CAGCTAGCG GACCTCTCG GTAGTGCGA CAAACTGGA GGTATCTTCT GTGCGCTCG CTAGCTGGA GCGCGCGGCC CTTGCCACGT

701 TTGGAACGG GATTCCCGT GCCAAGATG ACGTAGTAC GCGCTATAGA GTCTATAGCC CCACTTGGCT TCGTTAGAC GCGCTACAA TTAATACATA
 AACTTTCG CTAAAGGCA CATTCTCAC TGCATTATG CCGATATCT CAGATATCG GTGMAACGA AGCAATCTTG CCGCGATGT AATTATGTAT

15/35

FIGURE 5B

001 ACCTTATGTA TCATACACAT AGGATTTAGG TCACACTATA GAATACATC CACTTTGGCT TTCTCTCCAC AGGTOTCCAC TCCCAGGTCC AACTGCACCT
TGGATACAT AGTATGTGTA TGCTAATCC ACTGTGATAT CTTATTGTAG GTGAAACCGA AAGAGAGGTG TCCACAGGTG AGGOTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATTC CCCGGGATC CTCATAGAT CCGTCGACCT CGATCGACT TTTTGTAGG CCAAGGGTA CTCTTTTTTC
GCGAGATAG CTAACTTAAG GGGCCCTAG GAGATCTCTA GCGAGCTGA GCTCAGCTG AAAAAAANA AAAAAATCC GGTTCCTCAT GAAGAAAAAG

1001 TTTATTAAAT ACTCAGAAAT CTAGGCCACA GCAATCTACT GTTCTCCTCT CATTTCCTA AACTATTTG ATACCTATTT CTCAGACTTT ATGGGCTATT
AAATAATTAA TGAGTCTTCA GATCCGCTGT CGTTAGATGA CAGAGGAGA GTAAAGGAT TTGATAAATC TATGATANA GATCTGTANA TACCCGATNA

1101 AGACATTCT CACATTTCCA TAGATAATAA CTCATCCGTT TTGCAACCTG ATTCTCAATA TTAAGAGATT AAACTAATG TATATGACTC TCAGTTGACA
TCTGTAAAGA GTGTAAAGGT ATCTATTATT GAGTAGTCAA ACGTTGGAC TAAGAGTTAT ATTCTCTAA TTTTGATTAC ATATACTGAG AGTCMACTGT

1201 CATACTGAAG TACAGNAANA TTCCATCAAT TCCTTCTGCA AAATGMAAA GACTTCGTT TCTCAACAGC TGCAATATTT TTTTATGCAT AGAAANAAT
GTATGACTTC ATGCTCTTTT AAGGTAGTAA AGGAGAAGT TTTACTTTTT CTGAGGCAAA AGAGTTGTG ACCTAGTANA AAATACGTA TCTTTTTTTA

1301 GTGCAATTAC TCCAAATACA ATCAAGTCAT TTAACATGGC TTTACCATCA TTGTAGTTAC AGGATATTT AAAGAGANA AAAAAATCTC AAGGCACAGG
CACGTTAATG AGGTTCAATG TAGTTCAGTA AATTGTACCG AATGGTAGT ACATCAATG TCTATANA TTTTCTCTTT TTTTATAGG TTTCTGTCC

1401 TCCTGCTGTG CAGCAAGCA ATCAAAATCC TTCAATAA CAGCTGATG GGATTCAGCA ATCTGAAGAA TAATGAATA CCACCTAAT CAGTAAACAG
AGGACACAC GTGTTTCTGT TAGTTTAAGG AAGTATTATT GTGGGACTAC CCTAAGTCT TAGACTCCTT ATTACTTATT GGTGAGATTA GTCATTTGTC

1501 GAAATTTCTA CAACAGTCAC TGAGTAAAAA TTGCACTATC ATCTGTGAT TCTCTGATC GACATTTCAA ACAATAATG GAATGTAG TATCTCTTAA
CTTTTACCAT GTGTCAATG ACTCATTTTT AACTGTATAG TAGACAACTA AGAAGACTAG CTGTAAAGT TGTATTATTC CTTTACATTC ATAGGAAAT

16/35

FIGURE 5C

1601 AAGAAATAAT AACTTGCTTT AGTGTGCTTA ATTTTACAG GCACTGAGGA AATTATATAT CACCTTGACT GTCTGCACT GTTGCCCACT CAATAAATG
TTTCTTTTAA TTGAACCAA TCACACGAAT TAAATATGTC CGTCACTCCT TTAATATATA GTGAACTGA CAGGACGTCA CAACGGGTCA GTTATTTTAC

1701 CACAAATAAT CTTTTCATATA ATACATGGCC AACTTTATTC TATCACTTGA ATATGTGAGG ATAACTGAT TGTGCACTTG GTTGATAACA TTGTATTTTG
GTGTTTATTA GAAAAAGTAT TATGTACCGG TTGAATATAG ATAGTGAAT TATACAGTCC TATTGACTA ACAGGTCAAC CAACCTATTGT AACATNAAC

1801 GAATGATTA TTGAAATTTG TTTTGCTACT TTATTATTG ATATTCTTCT CCAATGTTC TCTTATGAA TTTATTCAT CTGAATATGA AGAGTCTGTT
CTTACCTAAT AACTTNAAC AATCGATGA AATATNAAC TATAGAAGA GGTCACAAGT AGAATACTTC AATAACGTA GACTTATCT TCTCAGACAA
O R I P N N A D S Y S S D T
506

1901 TCAAAATAGT CTTCAAGTTT CCAAGGCACT GTCTCAATG TAGTGCTTC CTTAGGCTCT GCATTCCAGC ACTCCAACAT GATGTGTAA AATGCTGTG
AGTTTATCA GAAGTTCAAA GGTGCGTCA CAGAGTTTAC ATCCAGCAAG GAATCCGAGA CGTAAGGTG TGAGTTGTA CTACAACATT TTAAGCACAC
492 E F Y D E L K W R L T E F T P R E K P E A N M C E L H I N Y F Q Q

2001 CACAGTTGGA TGGTTGCGA AGTCTATAGT TTTGAGCCAA CATCTGANT ACCTGGCAC CTGTATACC ACTGTAGGC ATTTGCCAT AGTAAATGAT
CTGTCAACCT ACCAAGCCT TCAGATATCA AACTCGGT GTAGACCTAA TGGACCGTG GACAGTATG TGACATCCG TAAACCGTA TTCATTACTA
459 P C N S P Q P L R Y N Q A L H Q I V Q A G T H G S Y P H K G Y T I I

2101 TTCAATAAGA AGGATTCCAA ATGACCATAC ATCGGACTTA ATGCTGAATT TATTACTAG ANTGCTTCG GGGCAGTCC ACTTCACCG CACCTTTATT
AAGTATTTCT TCCTAAGGTT TACTGTATG TACCTGAAT TACGACTTAA ATATGATGC TTACGGAAGC CCGGTGAGG TGAAGTGGC GTCGAATATA
425 E Y L L I G F S W V D S K I S F K N S R I A E P A T W K V P L K I

2201 TCGTGTAG ATTCAATAGT GTCTTCATTA TCTACTTAA AACTCTGGC AAGTCCAAA TCTGCTACTT TGTAGATATT ATGTCACCA ACGAGACAT
AGCACAGATC TAAGTATCTA CAGAAATAT AGATGAATT TTTGAGACCG TTAGGTTTT AGAGATGAA ACATCTATAA TACAAGTGT TCTCTCTGA
392 E H R S E Y I D E N D V K F V R A L G F D A V K Y I N H E G V L V

2301 TTCTGGCAGC CAGATCTCTG TGAATGATG TCCGACATC CAGATAGGCC ATTCCAGAGG CAACTGTGC CCGCATGTCT ACCTGTTGAG TCAGATGAT
AAGACCGTGC GTCTAGAGAC ACTTACATCA AGGCTCTGAG GTCTATCCG TAAAGTCTCC GTTGACACG CCGGTACAGA TGGACAATC AGTCTACCTA
359 N R A A L D R H I Y N R S E L Y A H Q S A V Q A A H D V Q Q T L H I

17/35

FIGURE 5D

2401 TTTTGATCCA GTGTCAATTT GGAGATATTC TTGCAGACTT CCATGTCTCA TCAACTCTGT AATAATATAA ATTGATCTT CTAAAGTCCA AACAGCATAA
 ANAAGTAGGT CACAGTAAAC CCTCTATAG AAGTCTGNA GGTACAGAGT AGTTGAGACA TTATTATATT TAACCTAGAA GATTTCAGT TTGCTGTATT
 325 K S G T D N Q L Y E Q L S G H R M L E T I I Y I P D E L T C V A Y

2501 AGCTGATAA GCTTTGGATG TCTTAGGTTT TCCATTATCT GTCCCTCCCT CAGGAGTCA TTGGATCCA TTGAACCTGG TTTTAATGTT TTCACTGCTA
 TCGACCTATT CGAACTTAC AGAATCCAG AGTAATAGA CACGGAGGA GTCTTCACT AACTTAGGT AACTTAGAC ANAATTACAA AAGTGACGAT
 292 L Q I L K P H R L N K H I Q A E R L F D N P D H S Q P K L T K V A

2601 CTGAGCTGTT ATTGTTCCAC AGACTTCCC ATACTTCGCC AACTGACCA GATCCCAATC CTTTCAGAG CTGTATGGAG TTGGGTCTA TCTCCCATG
 GACCTCACCA TAACAAGGTG TCTGGNAGGG TATCAAGCGG TTGACTGCTT CTAGGGTTAG CGAAGTCTTC GACATACCTC AACGCCAGAT AGAGGGTAAC
 259 V P T T N N W L G E W V E G F Q G S G L R K L L Q I S N R D I E W Q

2701 GTCCACGGT TTATACGACA ATCAAAATGG AGCTGAGACC TGATCTTTA AGCATGGTTT CCCAGCTTG ACACACAGC CGTCACTGTT CTGGGTGAG
 CAGGTGCCAA NATATGCTGT TTAGTTTACC TCGACCTGAG ACTAGAAAT TCGTACCAGG GGGTTCGAG TGTGTGTCG GCAGTGAACA GAACCAATC
 225 D V T K Y S L D F P A P V Q I K L C P K G L K V C L G D S T K T Y

2801 TGGCTCACAA ATTCTGTCAG TGTGAAAG ATTCTCTTTC GGTGAGAAA AAATCCCTCT TCATCCAGTC TTTTAATCT GTAGTGTTT ACAACTGCTC
 ACCGAGTGT TACGAGTC ACAACTTTTC TAAGAGAG CGCACTCTTT TTAGGGGGA AGTAGGTCAG AAATTAAGA CATCACAAA TGTGACGAG
 192 H S V F E N L T S F I R R R T L F F G Q E D L R K I R Y H K V V A

2901 CATCTAAAC TGAAGAGAG AATCTCTCTT TTGCTCTTTC ACTTCTCTG ATTAGAAAG AACCGTCTT GTTTCTGAA TATAATAGTT GTTCTCTGC
 GTAGATTTTG ACTTCTCTC TTAGAGGANA AAACCGAAG TGAAGAGAC TAATCTTTCC TTGGCCAGAA CAAAGACTT ATATTATCA CAAAGAGAG
 159 G D L V S L S F E G K Q S E S E R I L F S G T K N E S Y L L Q K E A

3001 ATCTGATCTT CCGATTGCTC CAAAGAACCA CGCTCTGCCC TGTAGGCTTC TGTCTCAGC CAGTAGTTA GAAGAAAT AGCCTGTAG TTGCTGACTG
 TAGACTAGAA GGCTAACGAG GTTCTCTGTT GCGGAGCGG ACATCCGAG ACAGAGTGG GTGCATCAAT CTCTCTTATA TCGGAACATC AACGACTGAC
 125 D S R G I A G F F W P E A Q L S R D E A V Y N S P I Y G Q L Q Q S

3101 GAGCCATCTC GTCTTTTCTC CAAAGTCTG GCAACCAACC AGCCCTCATG CAAAGTCTC AGAAGTCTC GTTTGTCACC TGTCTGGAAG CTCGAAGTCTC
 CTCGGTAGAG CAGAAAGAG GTTCACAGAC CGTTTGTGCG TCGGAGTAC GTTTGACAGG TCTTGAAGCTT CAAACAGTGG ACCAGCCTTC GAGTTGAGGA
 92 S G D R R K E L H R A F W W G E H L T D L V Q L K D G A R F S L D

18/35

FIGURE 5E

3201 CAGCAGTCCG AGCTGGTAA TCACAAAG CCACAAAGTA GTGGCCATGC CTCTGTGACT GGGGAGGCA AAGGCCCCCT GGATTTTCAA TCACGGTTGA
 GTCGTCAGGC TCGACCAAT AGTTGTTTC GGTGTTTCAT CACCGGTAGC GAGACACTGA CCCCTCTCGT TTCCCGGGA CTTAAAGTT AGTCCCACT
 59 E A T R A Q Y D F L A V F Y H G H R Q S Q P S C L A G P N E I V T S

3301 CTTCTCTGCC TCCGTGGACA AACAGGGGAG ATAGGTTTCT AGTACTCCC AGAGCCTCTG ACAGATGTTG CTCATTGTGC CTTCGTGGGG AGACAGGAG
 GAACAGAGCG AGCACCTGT TTGTCCCTC TATCCNAGA TCCATGAGG TCTCGGAGAC TGCTACACAC GAGTAACACG GAACCAACCC TCTTCTCTC
 25 K D A E T S L C P L Y P E L Y E W L R Q C I N S M

3401 CAGGCTTCT CCCTCTCCC TTAGTCTCTG CGATCCACCT TATCTTCTCT CACCAGGCA CTTTGAAGTC AGCACCACT CACCATACCT CGAGAGTAT
 GTCCCGNAGA GGGAGAGGGG ATCAGAGAC GCTAGGTGGA ATAGAGGMA GTGTCCCTT GAACTTTCAG TCGTGGTTGA GTGATATGAA GCCTCTCATA

3501 GCAAAGTCCC GTTTCAGATC AGTCCAGCAG CTGGGTTGCA GCAAGTCTTA CCTGGAGAGA CTTACCGCT TCTTCTCTGT GGCTGGAGGT GCTACCCCGA
 CGTTTCAGGG CAAGTCTAG TCAAGTCTGTC GACCCACGT CGTTCAGAT GGACTCTCT GAAATGGCGA ACNAAAGACA CCGACCTCCA CGATCGGCT

3601 GGCNAACTG AGCAGAGCT GGGCAGCTGC TCACATAGAA GGTGCTTTT CTCTTATCT CTTAAGNAT CCCACNACA AATATAATA AATATAAAG
 CCGTTTTGAC TCGTCTCTGA CCCGTCTCGA AGTATCTCT CCACAGAAA GAAGATAGA CGAATCTTA GGGTGTGTT TTAATTTTAT TTAATTTTTC

3701 GGCTTTATTT AGACAAATAT CTGAGACAG AATGATGCA TCTTGCTTT TGTCCATA AAGGTTAGC AAGAGGAGC TACTAACCC TGGTAAACCC
 CCGAATATA TCTGTTTATA GACTCTGTC TTACCACGT AGACGGAA ACAGGGTTAT TTTTCATCG TTCTCTTTCG ATGATTCGG ACCATTTTCG

3801 TCCACGCTT GCTTTGCCA GGTTCGACTC GAGATATCTT CCATACCTAC CAGTCTCTCG CCTGCAGTTC GGGCCCGCA CTCTAGAGTC GACCTGCAGA
 AGGTGCAGAA CGAAGCGGT CCCAGCTGAG CTCCTTAGA GGTATGGATG GTCAGACGC GAGCTCCAG CCGCGGCGCT GAGATCTCAG CTGGACGCTCT

3901 AGCTTGCCG CCATGGCCA ACTTGTTTAT TCGAGCTTAT AATGGTTACA AATAAGCAA TAGCATCACA AATTCACAA ATAAAGCAT TTTTTCACATG
 TCGAACCGG GGTACCGGT TGAACATA ACGTGAATA TTACCAATGT TTATTTCTGT TTAAGTGT TATTTCGT AAAAGTAC

19/35

FIGURE 5F

4001 CATTCTAGTT GTGGTTTGT CAACTCATC AATGTATCTT ATCATGTCTG GATCGGGMAT TAATTCGGCG CAGCACCATG GCCTGAAATA ACCTCTGAAA
 GTAAGATCA CACCAACAG GTTGAGTAG TTACATAGTA TAGTACAGAC CTAGCCCTTA ATTAAGCCGC GTCTGGTAC CGGACTTTAT TCGAGACTTT

4101 GAGGAACCTG GTTAGGTACC TTCTGAGCG GAAAGAACCA GCTGTGGAAT GTGTGTCAGT TAGGGTGTG AAGTCCCA GGTCTCCCG CAGGCAGNAG
 CTCCTTGAC CAATCCANTG AAGACTCCG CTTTCTTGGT CGACACCTTA CACACAGTCA ATCCACACC TTTCAGGGGT CCGAGGGGTC GTCCGTCTTC

4201 TATGCAAGC ATGCATCTCA ATTAGTCAGC AACCAAGTGT GGAAGTCCC CAGGCTCCC AGCAGGCAGA AGTATGCATA GCATGCATCT CAATTAGTCA
 ATACGTTTTC TACGTAGAGT TAATCAGTCT TTGGTCCACA CCTTTCAGCG GTCCGAGGG TCGTCCGTCT TCATACGTTT CGTACGTAGA GTTAAATCAGT

4301 GCAACCATAG TCCGCCCCCT AACTCCGCC TAACTCCGCC CAGTCCGCC CATCTCCGC CCATGGCTG ACTAATTTTT TTTATTTATG
 CGTTGGTATC AGGCGGGGA TTGAGGCGG TAGGGCGGG ATTGAGCGG GTCAAGCGG GTAGACCGG GGTACCGAC TGATTAATAA AATAAATATC

4401 CAGAGGCGGA GGCGGCTCG GCCTCTGAGC TATTCAGAA GTAGTGAGGA GGCTTTTTG GAGGCTAGG CTTTGCATA AAGCTGTAA CAGCTTGCA
 GTCTCGGCT CCGGCGGAGC CGGAGACTCG ATAACTCTT CATCACTCTT CCGAATAAC CTCGGATCC GAATACGTTT TTCGACATTT GTCGAACCGT

4501 CTGCGCGTCTG TTTTACAAG TCGTGACTCG GAAAGCCCTG GCGTTACCCA ACTTAATCGC CTTCGAGCAC ATCCCTTTT CCGCAGCTG CGTAATAGCG
 GACCGGCAGC AAATGTTTC AGCACTGACC CTTTTCGAC CGCAATGGT TGATTTAGCG GATCTGCTG TAGGGGAAA CCGGTCCACC GCATTATCGC

4601 AACAGGCCCG CACGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGCG GAATGGCGC TGATGGGTA TTTCTCCTT ACGCATCTGT GCGGTATTTT
 TTCTCCGGC GTGGCTAGC GGAAGGTTG TCACCTGTC GCACTTACCG CTTACCGCG ACTACGCCAT AAAGAGGAA TCGGTAGACA CGCCTAATAG

4701 ACACCCGATA CGTCANAGCA ACCATAGTAC GCGCCCTGTA GCGGCGCATT AAGCGCGCG GTGTGTGTG TTACGCGCAG CGTGACCGCT ACCTTGCCA
 TGTGCGGTAT GCATTTCTGT TGGTATCATG CCGCGGACAT CGCCGCGTA TTGCGCGCG CACACACACC AATGCGGTC GCACTGCGCA TGTGAACGGT

20/35

FIGURE 5G

4801 GCGCCCTAGC GCGCGCTCCT TTGCTTTCT TCGCTTCTT TCGGCCACG TTGCGCGGCT TTGCGCGCTCA AGCTCTAAT CGGGGGCTCC CTTTAGGGTT
CGCGGANTCG CGGCGAGGA AGCGNAGA AGCGNAGGA AGAGCGGTGC AAGCGCGCA AAGGCGCAGT TCGAGATTTA GCGCGCGAGG GAAATCCCAA

4901 CCGATTTAGT GCTTTACGGC ACCTCGACCC CAANAACTT GATTGGGTG ATGTTTCACG TAGTGGCCA TCGCCCTCAT AGACGGTTTT TCGCCCTTTG
GGCTAAATCA CGAATGCCG TCGAGCTGG GTTTTTGAA CTAAACCCAC TACCAAGTGC ATCACCCGAT ATCGGACTA TCTGCCAATA AGCGGGAAC

5001 ACGTTGGAGT CCACGTTCTT TAATAGTGA CTCTGTTCC AACTGGAAC AACACTCAC CCTATCTCG GCTATCTTT TGAATTTATA GGAATTTTGC
TGCNACCTCA GGTGCAGAA ATTATCAGCT GAGMCAGG TTGACCTTG TGTGAGTTG GATAGAGCC GATNAGAA ACTAATATT CCTTANAAACG

5101 CGATTTCCGC CTATTGGTTA AAAATGAGC TGATTTNACA AAATTTNAC GCGAATTTA ACAAAATAT AACGTTTACA ATTTATGCT GCACTCTCAG
GCTAAGCCG GATAACCAAT TTTTACTCG ACTAATTTGT TTTTAAATG CGCTTAAAT TGTTTTATA TTGCAATGT TAAATACCA CGTGAGAGTC

5201 TACAACTGC TCTGATGCCG CATAGTTAG CAGCCGCCA CACCCGCTGA CCGCCCTGA CCGGCTTGC TGCTCCCGG ATCGGCTTAC
ATGTTAGAC AGACTACGC GTATCAATTC GTGCGGCT GTGCGGACT GTGCGGACT GCGCGGACT GCGCGGACT GCGCGGACT TAGCGGATG

5301 AGACAGCTG TGACCGTCTC CGGAGCTGC ATGTGTCAGA GCTTTTCACC GTATCACCG AACCGCGCA GACGAAAGG CCTCGTGATA CGCTATTTT
TCTGTTGAC ACTGCGAGG GCGCTCGAG TACACAGTCT CCAAAAGTGG CAGTAGTGG TTGCGCGCT CTGCTTTCCC GGAGCACTAT GCGGATAAAA

5401 TATAGGTTAA TGTGATGATA ATAAATGTTT CTAGAGCTC AGGTGGCACT TTTCGGGAA ATGTGCGCG AACCCCTATT TGTATTATTT TCTAATACA
ATATCCAAAT ACAGTACTAT TATTACCNA GAATCTGCAG TCCACCGTGA AAGCCCTT TACACCGCC TTGGBGATA ACAANTAAA AGATTTATCT

5501 TTCNAATAG TATCCGCTCA TGAGACAATA ACCCTGATA ATGCTTCAAT ATATTTGAA AAGGAGAGT ATGATTTTC AACATTTCCG TGTCGCCCTT
AAGTTATAC ATAGCGAGT ACTCTGTTAT TGGACTATT TACGAGTTA TTATACTTT TTCTTCTCA TACTCATAG TTGTAAAGC ACAGCGGAA

21/35

FIGURE 5H

5601 ATTCCCTTTT TTGCGGCATT TTGCTTCTCT GTTTTTCCTC ACCCAGAAAC GCTGCTGAAA GTAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT
TMAGGGAAAA MACGCCGTAA MACCGAAGGA CAAMAGCAG TGGGTCTTTG CGACCACTTT CATTTTCTAC GACTTCTAGT CAACCCACGT GCTCACCCCA

5701 ACATCGAACT GATCTCMAC AGCGGTAAAG TCCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC
TGTAGCTTGA CCTAGAGTTG TCGCCATTCT AGGAACCTCTC AAMAGCGGGG CTTCCTTCCAA AAGTTACTA CTCGTGAAA TTTCAGAGCG ATACACCGCG

5801 GGTATTATCC CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GATGACTTG GTTAGTACT CACCAGTCAC AGAAAGCAT
CCATAATAGG GCATAACTGC GCGCCGTTCT CGTTGAGCCA GCGCGGTATG TATTAAGAT CTTACTGAAC CAACTCATGA GTGGTCAGTG TCTTTTCGTA

5901 CTTACGGATG GCATGACAGT AAGAGAAITA TCCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGCCA ACTTACTTCT GACAAAGATC GAGGACCGA
GAATGCCCTAC CGTACTGTCA TTCTCTTAT AGTCACGAC GGTATTGGA CTCACTATTG TGACCGCGT TGAATGAAAG CTGTTGCTAG CCTCTCGCT

6001 AGGAGCTAAC CGCTTTTTG CACACATCG GGCATCATGT AACTGCGCTT GATCGTTGGG AACCGAGCT GAATGAGCC ATACCAAAAG AGAGCGTGA
TCTCGATTG GCGAAANAC GTGTGTACC CCTAGTACA TTGAGCGGA CTAGCAACC TTGCGCTGA CTACTTCGG TATGGTTTGC TGCTCGCACT

6101 CACCACGATG CCTGTAGCAA TGCCAAACAC GTTGGCGAA CTATTAACTG GCGAACTACT TACTTAGCT TCCCGGCAC AATTAATAGA CTGATGAG
GTGGTCTAC GGACATCGTT ACCGTGTG ACCCGGTTT GATMAATTGAC CGCTTGATGA ATGAGATGGA AGGCGCGTTG TTAATTATCT GACCTACCTC

6201 GCGGATTAAG TTGACGACC ACTTCTGCG TCGGCCCTTC CGGTGGCTG GTTTATTGCT GATMAATCTG GAGCCGTTGA GCGTGGTCT CCGGTATCA
CGCCTATTTC AACGTCTTG TGAAGACGG AGCCGGMAG GCGGACCGAC CAATAACGA CTATTAGAC CTGCGCACT CCGACCCAGA GCGCCATAGT

6301 TTGCAGCACT GCGGCCAGAT GGTAAACCT CCCGTATCGT AGTTATCTAC ACCAGCGGA GTGAGCGAAC TATGATGAA CGAAATAGAC AGATCGCTGA
AACGTCTGA CCGCGTCTA CCATTGCGA GCGCATAGCA TCAATAGATG TGCTGCGCT CAGTCCGTTG ATACCTACTT GCTTTATCTG TCTAGCGACT

22/35

FIGURE 5I

6401 GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAACTTTTACT CATATATACT TTAGATTGAT TTAATACTTC ATTTTAAAT TAAAGGATC
CTATCCACGG AGTCACTAAT TCGTAACCAT TGACAGTCTG TGACAGTCTG GTTCAAAATGA GTATATATGA AATCTAACTA AATTTTGAG TAAAAATTAA ATTTTCTCTAG

6501 TAGGTGAAGA TCCTTTTGA TAACTCATG ACCAAATCC CTTAACGGTA GTTTTCGTTT CACTGAGCGT CAGACCCCGT AGAAAGATC AAGGATCTT
ATCCACTTCT AGGAAAACT ATTAGAGTAC TGGTTTTAGG GAAATTGCACT CAAAGCAAG GTGACTGCA GTCTGCGCA TCTTTCTAG TTCTCTAGAA

6601 CTTGAGATCC TTTTCTCTG CCGGTAACT GCTGCTTGA AAAAAAAA CCACCCCTAC CAGCGGTGCT TTGTTTCCG GATCAAGAGC TACCAACTCT
GAACTCTAGG AAAAAAGAC GCGCATTTAGA CGACGAACGT TTGTTTTTTT GGTGCGCATG GTCCGACCA AACAAACGGC CTAGTTCTCG ATGTTTGAGA

6701 TTTTCCGAAG GTAACTGGCT TCAGCAGAGC GCAGATACCA AATACTGTTT TTCTAGTTTA GCGGTAGTTA GCGCACCACT TCAAGAACTC TGAGCACCG
AAAGGCTTC CATTCACCGA AGTGGTCTCG CGTCTATGCT TTATGACAAAG AAGATCACAT CCGCATCAAT CCGGTGTTGA AGTTCTTGAG ACATCGTGGC

6801 CCTACATACC TCCTCTGCT AATCTGTGA CCAGTGGCTG CTGCCAGTGG CGTAAAGTGG TGTTCTTACCG GGTGGAAGT AAGACGATAG TTACCGGATA
GGATCTATCG AGCGAGACGA TTAGGACAAAT GGTCAACCGAC GACGGTCACC GCTATTTCAGC ACAGATGCG CCAACTGAG TTCTGCTATC AATGGCCTAT

6901 AGCGGCACCG GTGCGCTGA ACGGGGGTT CGTGACACA GCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAATAC CTACAGCGTG AGCTATGACA
TCCGCGTGGC CAGCGGACT TGCCCCCNA GCACGTGTGT CCGGTGMACT CTGGCTTGT GATGTGCT TGACTCTATG GATGTGCGAC TCGATACTCT

7001 AAGCGGCACG CTTCCCGAAG GCAGAAAGGC GCACAGGTAT CCGGTAAAGC GCAGGTCTCG AACAGAGAG GCGACGAGGG AGCTTCCAGG GCGAAACGCC
TTCCGCGTGC GAAGGGCTTC CTTCTTTCCG CTTGTCCATA GCGCATTCGC CGTCCGAGCC TTGTCTCTC GCGTGTCTCC TCGAAGGTCC CCTTTTCCG

7101 TGGTATCTTT ATAGTCTGT CCGGTTTCCG CACCTCTGAC TTGAGCGTGG TTGTTTGTGA TGCTGCTCA GCGGCGGAG CCTATGGA AAGCGCAGCA
ACCATAGAAA TATCAGGACA GCGCAAGCG GTGGAGACTG AACTCGCAGC TAAAAACT ACAGACAGTC CCCCCCCTC GGTATCTTT TTGCGGTCTG

23/35

FIGURE 5J

7201 ACGCGGCTT TTTACGGTTC CTGGCCTTT GCTGGCCTTT TGCTCACATG TTCTTTCCCTG CGTTATCCCC TGATTCTGTG GATAACCTA TTACCGCCTT
TGCGCCGGAA AATGCCAAG GACCGGAAA CGACCGGAAA ACGAGTGTAC AGNAAGGAC GCNATAGGGG ACTAAGACAC CTATTGGCAT AATGGCGAA

7301 TGAGTCAGCT GATACCGCTC GCGCAGCCG AACGACGAG CGCAGCGAGT CAGTGAGCGA GGAAAGCGGA GAGCGCCCA TAGCNAACC GCCTCTCCCC
ACTCACTCGA CTATGGCGAG CGGCTCGGC TTGCTGGCTC GCTGCTCTCA CTCACTCGCT CCTTGGCCTT CTGGCGGTTT ATGCGTTTGG CGGAGAGGGG

7401 GCGCGTTGGC CGATTCATTG ATGCAGCTGG CACGACAGGT TTCCGACTG GAAAGCGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT
CGCGCACCG GCTAAGTAAT TACGTGACC GTGCTGTCCA AAGGGCTGAC CTTTGGCCCG TCACTCGCT TCGTTAATT ACACTCAATC GAGTGAGTAA

7501 AGGCACCCCA GCGTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTAGCG GATAACAAAT TCACACAGGA AACAGCTATG ACATGATTAC
TCCGTGGGT CCGAATGTG AATACGAAG GCCGAGCATA CAACACACT TAACACTCGC CTATTGTTAA AGTGTGCTT TGTGATAC TGTACTAATG

7601 GAATTAA
CTTAATT

24/35

1 GCGGCCGCAG AGAAGCAGA GGATGGGGCT TACGAGCTGG CAGAGCCAGG AGCGGGGAGG TAGCAGAAAG ACCACAAGTA CAAAGAAGTC CTGAACATTT
CGCCGGCGTC TCATTGCTCT CCTACCCCGA ATCGTCGACC GTCTCGGTCC TCGCCCTCC ATCGTCTTC TGGTGTTCAT GTTCTTCAG GACTTTGAAA

101 GGTTTTGCTG CTGCAGCCCA TTGAGAGTGA CGACATGGAG CACAAGACCC TGAAGATCAC CGACTTTGGC CTGGCCCGAG AGTGGCACAA AACCACACAA
CCAAACGAC GACGTGGGT AACTCTCACT GCTGTACCTC GTGTCTGGG ACTTCTAGTG GCTGAACCG GACCGGGCTC TCACCGTGT TTGGTGTGT

201 ATGAGTCCCG CAGGCACCTA CXCTGGATG GCTCCTGAGG TTATCAAGGC CTCCACCTTC TCTAAGGCA GTGACGTCTG GAGTTTGGG GTGCTGCTGT
TACTACGGC GTCCTGGAT GTGNCCTAC CGAGGACTCC ATAGTTCCG GAGGTGGAAG AGATTCCCGT CACTGCAGAC CTCAAAACCC CACGACGACA

301 GGGAACTGCT GACCGGGAG XTGCCATACC GTGGCATTGA CTGCCCTTGGT GTGGCCTATG GCGTAGCTGT TAACAGCTC ACAGTCCAT CCATCCACCT
CCCTTGACGA CTGGCCCTC PACGTATGG CACGTAACCT CACGGAACGA CACCGGATAC CGCATCGACA ATTGTTCGAG TGTGACGTA GTTAGCTGGA

401 GGGC
CCGG

FIGURE 6

25/35

FIGURE 7A

1 ATGACAGGCT TGGGCGGCGA GCGCGGCGAG CTGGCGGCTG TGGCTGCTTT TTGTTGAAATG ATATTGCGA GTATTACAAA TCAAGATATG GCTGTGATCA
TACTCTGGA ACCGCGGCTT GCGGCGGCTG GAGCGGCGAG AGCAACAAAA AGACGCTTAC TATAAGAGCT GATAATGTTT AATTCTAAG AGACACTAGT

101 AATGCTGTTT AATCAATAT AAGAACAAATG ATTATCAAT GCGGAAGTGA TGTATATAT GCATGCTATC AGAATCCCGG GAGACGCTG GGTGTGCTTT
TGACACAAAA TTATTAATA TTGTTGTTAG TAATTAAGTGA GCGCTTCAAT AGTATATAG GTATGCTAG CTCTGAGG CTCTGAGG CTACACGCAA

201 GAGACGCGAG AGCTCAGGGA GAGTGTACGA AGTGTGCTT GTCGAAGTGG ATGTATGTC TTGATATGCA CTGCAAGTGC TGTGCTGATCC CCCAGGGAAC
CTCTGCGCTC TGAAGTCTCT GTACATGCT TGCACGCTGA GAGCTTCAAG TACATAGAG AGCTGAGTGT GAGCTTCAAG ACCAGCTAGG GGTGCTCTTG

301 ATTTCCTGTC TGTGCTCT TAAGCACAG TGTGTAAT GCGAGTACA TTTGATTTA GAAACACAG GAGTGTGTTT CATGCTCAT TTGAATATGA
TAAGGATAG AGACCCAGAA ATTGCTGTC AGGACCTTAA GCGCTGCTT AAGACTAAAT GTTTGCTGTC GTACACAAAG GTACCAATAA AACTTTTACT

401 GAGAAAGCA AGTGTGAGAA TACCTAATTT TTATTGAG TGAAGTATC AATTACAAA TATTGTTTAC AGTGAATATA AGAATATGCC TTTTTTACAC
CTCTTGTCT TCACTCTT ATGATGAAA AATAATGTC ATTGCTGTC GCGACGACA GTATAGAT CTGCGAAGT CTGCGCTAGG ACCTTACGCA GGAACGCTA

501 ATTAAGAGA CTTACTTTA GAAATATGA AAGCAGGAG GCGCTGCTT GCATATCTGA GAGCTTCCA GAGCGATCC TGGAAATGGT GCTTGGCAT
TAATCTCTT GCAATGAAT CTTTCTAAT TTTGCTCTG GCGACGACA

601 TCAAGCGG AGAGCTTAA AGACAAAT GCGCTGCTT TTAAAGGA GGAAGATG CTTCAATAT TATTGGGAC GGACATAGAG TGTGTGCGCA
AGTGTGTC TTGACAT TCTTCTTTA TCTTCTTTA GCTCAGAG AATTTCTT CTTTTTAC GAGTACTTA ATAAAGCTG CCTGTATTC ACACACGCT

701 GAAATGAAT GCGAGGGA TGAAGAGG TTTGCAAT AGATTAAAT CAACTGTC AGACACAT GCGACAAAT TTTCTAAG TACGGAAC
CTTACTGA GCTGCTT ATGCTGTC AGAATGTTA TGTACATTA GCTTGAAG TCTGTATA GGTGTTAAT AAGAATTT ATCTGTTG

26/35

FIGURE 7B

801 CTTATGATATA AGGTGCAAAAG CTGTTGATAT GAAAGATGCA TTGAGGCTCA CCGGGAATT AAAAAACAAA GAACTCGAGG AGGGCACTA CTTGAGATG
 GAAATATAT TCCAGCTTC GACAACTACA CTGCTATCT AAGCGGAT GACCCCTTAA TCTTTTCT CTTGAGCTCC TCCCGTTAT GAACTCTAC

901 ACTAGCTATT CAACAAACAG AACTATGATA GCAATTTCT TGGTTTTCT ATATATCTG GCAAGAAAG ACACGGGATA CTACACTTGT TCGTCTTCAA
 TATGATATA GTTGCTTCTG TTTATATAT GCTATATAT GCTAAAGAA AAGCAAAACA TATATGCTAC GTTCTTTTC TGTGCGCTAT GATGTGAACA AGGAAGATT

1001 AGCATGCGAG TCAATCAAT TGGTTACCA TCGTACAAA GCGATTATATA AATGCTACCA ATTCACTCA AGATTATGAA ATTGACCAAT ATGAAGATT
 TCTATGCTC ACTTACTGCA AACCAATGTT AGCATCTTT GCTAAATAT TTACATGTT TAACTTACT TATAATATCT TAACTGTTA TACTTCTCAA

1101 TTTTTTTCT GTAGTTTA AAGCTAGC ACAAATGAA TTATGCTGCA GTTCTCTG GAAATCATTT CTTTGTGAG AAGAGTCT TGTAAACGGA
 AAAAAACA CATGCCAAT TCGATGAG TCTTACTT ACATGACT GAAAGAGAG TTTTGTGAAA GGAACACTG TTTTCCCAAC ACTATTGCT

1201 TAAAGCATAT CCAAGTTTTC CAATCATAG CACCAAGCAG GAGATATAT ATTGATGCA GAATGATG ATGCCAAT TACCAAAATG TTCACGCTGT
 ATGCTGATA GTTCAAAAC GTTAGTATC GTGCTGCTC GTTTATATA TAACTACT CTTTTACTAC TACGCTTAA ATGCTTTTAC AAGTGGCACA

1301 ATATAGAA GAACTCAA GTCTGCGAG AAGCTTGGC AATGAGCG TCGTGTCTT CGGATGATA CCAATTACCA TCTTGCACT GAAAGAAAGT
 TATATCTTC CTTGCAAT CAGAGCTC TCGAAGCG TTAACTGAG AHAACAAACA GCTACTAT GGTAAATGT AGAACCTGCA CTTTCTTAC

1401 TTCAAGCAAG TGTGCAACT GACAGAAACA GATACAAA GAACTGCTG AATAGAGAG TAACAGAAA GTTTTGGAC AGTGTGTC GAACTACT
 AATCTGTC AGAGGTTCA GTTGTCTT CTACTGCTT GTTATGCTT GTTATGCTT CACAAAGCTG TCAACCAAG CTGCTCATGA

1501 CTAAACATG GTGAGCCAT AAAAGCTC CTGTCAAT GTGTGCTA CAATTCCTT GGCATCTT GTGACAGAT CTTTTAAAC TCTCCAGGCC
 GATTCTAT CACTGCTA TTTTCCAG GAACTGCTA GAACTAT GTTAAGGAA CCGGTAGAA CACTCTGCTA GGAATTTG AGAGTCCGG

27/35

FIGURE 7C

1601 CCTTCTCTTT CATCCAGAC AACATCTAT TCTATGCAAG AATTCCTTT TCTCTCTCT TCACTCTCT TTAACCTCT CTAATTTCT ACAATGACAA
 GGAAGGAAA GTAGCTTCTT TTTAGACTA AATATGCTT TTAAGCAAA ACAGAGCAA AGTAACAGA AATTCGAC CATTAACAG TCTCATCTT

1701 AAGCAATT AGTATGAAA GCAATCTACA GATCTACAG GTACAGCAT CCTCAGATAA TGAGTACTT-TAGCTTATT, TCAGAGATAA-TUAATATCAT
 TTTCTTTAAA TGCATACTT GATCTATCT GTAGCACTT CACTCTCTTA GGAGCTATT ACTCATGAG ATGCAACTAA AGTCTCTT ACTTATACTA

✓ 1801 CTCAANTGG-AGTTTCAAG AGAAATTTA GAGTTTGGAA AGGTACTAG ATCAGCTCT TTGGCAAG TGTGACGC AACAGCTTAT GGAATTAGCA
 T-GAGTTTACC TCAGAGTTC TCTTTTAAAT CTCAAACTT TCCATGATCC TACTCCAGCA AAACCTTTT ACTACTTCC TTGCGAATA CCTTANTCT

1901 AAGCAAGAGT-CTCAANTCAG GTTACCTCA; AATGCTCAA; AGAAAGCA GACAGCTCT AAGAGAGGC ACTCATGTCA GAATCAAGA TGTATACCCC
 TTTGCTCTCA GAGTAGCTC CAATGGCAGT TTTACGACTT TCTTTTCTT CTCTCGAC TTTCTCTCC TGACTACAGT CTGAGTCT ACTACTGGT

2001 GCTCTGAGC-CACAGATAA TTGTGACCT-GCTGAGGCG TGACACTGT CAGGACCAAT TTACTTGAT TTTGAATCT-CTTCTATGG TCATCTCTG
 TCGATCTCTT AACACTTGA CGACCTGCG ACCTGTGACA GTCTGTGTA AATGAATAA AAATTTATGA CAACGATACC ACTAGACAG

2101 AACATCTAA-GAATGAAAG AGAAATTT-CACAGCACTT GACAGAGAT TTTCAGGAA CACAATTTCA GTTTTACC CACTTCCAA TCACATCCAA
 TTTATAGATT CTTCATTTT TCTTTTAAA GTTCTGCAA CTGTCTTGA AAGTTCTT GTGTAAGT CAANAATGG GTCAAGCTT AGTCTAGCTT

2201 ATTCCAGCAT-GCTGCTTCA AGAGAAGTTC-AGATACACC GCACTCGAT CAATCTCAG GGTTCATCG GAATTCATT CACTCTGAG ATCAATTTGA
 TAAAGTCTGA CGACCAAGT TCTCTCAAG TCTATGTGG CTTGAGCTT GTTTAGAGTC CCGAGTACC CTTAAGTAA GTGAGACTTC TACTTTAAT

2301 ATATGAAAG CAANAAGGC TCGAAGAGA GGAGGACTTG AATGTGCTTA CATTGAGA TCTCTTTGG TTGCTATC AGTTCCTCAA AGAATGGA
 TATATTTTG GTTTTTCCG ACCTCTCTT CTCTCTGAG TTACAGCAAT GTAACTTCT AGAAGAACG AAACGTATAG TTCAACGGT TCTTACCTT

FIGURE 7D

2401 TTTCTGCAAT TTAAGTCGTG TTTTCACAGA GACCTGCGCG CGAGGAACGT GCTTGTACCC CACGGGAAG TGGTGAAGAT ATGTGACTTT GGATTGGCTC
 AAGACCTTA AATTCAGCAC ACAAGTGCTCT CTGACCGGCG GGTCCCTTGA CGAACAGTGG GTGCCCTTC ACCACTTCTA TACACTGANA CTTAACCGAG

2501 CAGATATCAT GACTGATTC AACTATGTTG TCAGGGGCAA TCGCCGCTG CCGTAAAT GGATGCGCCG CGAAGCCCTG TTTGAAGCCA TCTACACCAT
 CTCTATGCTA CTCACTAAGG TTGATACAACT TCGCCCGCTT CGGGCAGAC GGACATTTTA CTTACCGGGG GCTTTCGGAC AACTTCCGT AGATGTGCTA

2601 TAAAGTGCAT GTCTGCTCAT ATGGAATATT ACTGTGGGA ATCTTCTCAC TTCTGTGAA TCCTTACCCT GGCATTCCGG TTGATGCTAA CTCTACAAA
 ATTCTCACTA CAGACCCAGTA TACCTTATAA TCACACCTT TACAAGATG AACACACCTT AGGAATGGA CCGTANGGCC AACTAGCATT GAAGATGTTT

2701 CTGATTCAAA ATGATTTAA AATGGATCAG CCATTTTATG CTACAGAA CAATATACAT ATATGCAAT CCGCTGCGC TTTGACTCA AGGAAACGGC
 GACTAAGTTT TACCTAAT TTACTAGTC GGTAATAATC GATCTCTTCT TTATATGTA TATTAGTTA GGACGACCG AACTGAGT TCCTTTCCCG

2801 CATCCTTCCC TAATTGACT TCGTTTTTAC GATCTCAGCT GGCAGATCCA GAAGAGGGA TGTGATGCG CCGTCTTCGG AATGCTCTCA
 GTAGGAGGG ATTAAGCTA AGCAAAATC CTACAGTGA CCGTCTAGCT CTCTCTGCT ACACCTACCG GCACAAAGCC TTACAGGAGT

2901 CACCTACCAA AACAGCCGAC CTTTCAGCAG AGAGATGGAT TTGGGCTAC TCTCTCCGA GGTCTAGGTC GAAGATTGCT AGAGGAACAA TTTAGTTTTA
 GTGGATGGT TTGTCCGCTG GAAAGTCGT TCTCTACCTA AACCCGATG AGAGGCGGT CCGAGTCCAG CTTCTAAGCA TCTCTCTGTT AATCAAAAT

3001 AGGACTTCAT CCGTCCACCT ATCCCTAACA CGCTGTAGAT TACCAACA AGGTTAATTT CATCACTAAA AGAAATCTA TTATCAACTG CTGCTTCACC
 TCTGTAACTA GGGAGGTGGA TAGGGATTGT CCGACATCTA ATGGTTTTGT TCAATTTAAA GTAGTCATTT TCTTTAGAT AATAGTTGAC GACCAAGTGG

3101 AGACTTTTCT CTAGAGAGCG
 TCTCAAAAGA CATCTCTGCG

FIGURE 8A

1 TCGACGCTCCA CCGCCCCAGG GAGAGTCAGA CATTGACGCG CAGAGCCCC CCAACTCAG TTGGATGCT ACCGAGTGA GGGGGGCCCA TGGAGCTCCG
 AGTCCAGGT GGGGGGTCC CTCACGTCT GACCTGCTTC GTCCCGGCG GTTTGAGTC AAGCTTAGGA TGGGTCACT CCGCCCGGT ACCTCGAGGC
 M E L R

101 GGTTCCTGCTC TCGTGGGCTT CCGTGGCGG AGTTTTCGAA GAGAGCTGCG TGACACAAA ATTGGAACT GCTGATCTGA AGTGGGTGAC ATTCCCTCAG
 CACAGAGCAG ACGACCGGAA GCAACCGGCG TTGAACCTT CTCTGACAG ACTTGTTT TAACTTTGA CGACTAGACT TCACCCACTG TTAGGGAGTC
 V I L C W A S L A A A I E E T L L N T K L E T A D L R W V T F P Q

201 GTGACGGCG AGTGGAGCGA ACTGAGCGC CTGATGAGG AACAGACAG GTGATCACC TACGAAGTGT GTGACGTGCA GGTGCCCCG GGCACGGCCC
 CACCTGCCCC TCACCTCTCT TCACTGCGG GACTACTGCT TTGCTGCTG GCACTGCTG ATGCTTCACA CACTGCAGT CGACAGCGCG CCGTCCCGG
 DV D G Q W E E L S G I D E E O H S V R T Y E V C D V Q R A P G Q A H

301 ACTGCGCTTG CACAGGTTGG GTCCACGCG GCGGTGCTT CCAGCTGTAC GGCACGCTG GCTTCACCAT GCTCGAGTGC CTGTCCTCTG CTGCGGCTGG
 TCACCTGAGC GTGTCGACC CAGGTTGCG CCGTTCGCA GTTCACATG CCGTGGGACG CGAGCTGTA CGAGCTCAG GACAGGAGCG GAGCCCGACC
 72 W L R T G W V P R R I A V H V Y A T L R F T M L E C L S L P R A G

401 GCGCTCCTGC AAGGAGACT TCACCTCTT CTACTATGAG AAGGATCGG ACACGGCCAC GGGCTCAGC CCAGCTGGA TGGAGACCC CTACATCAAG
 CCGGAGGAGC TTCTCTCTGA AGTGGCAGAA GATGATATC TGCTAGCGC TGTCGGGTG CCGGAGTGC GGTCCGAGCT ACCTCTTGGG GATGTAGTTC
 105 R S C K E T F T V F Y Y F S D A D T A T A L T P A M H E N P Y I K

501 GTGACACCG TGGCGCGGA GCATCTCACC CCGAAGCGC CTGGGCGGA GGCACCGCG AAGTGAATG TCAGAGCGCT GGTCTGGA CCGCTCAGCA
 CACCTGTGCC ACCGCGCT CGTAGAGTGG GCCTTCGGG GACCCGCTT CCGTGGCGG TTCCACTTAC AGTTCTGGA CGCAGACCT GCGGAGTGGT
 138 V D T V A A E H L T R K R P G A E A T O K V N V K T L R L O P L S K

601 AAGTCTTCTT CTACCTGCG TTCCAGACG AAGTCTCTG CATGCGCTG CTATCCCTG ACCTCTTCTA CAAAGTGC GCGCAGCTGA CTGTGAACT
 TGTGCTTAA GATGAGTGG AAGTCTCTG TGTGAGTAC GTACCGGAC GATAGTACG TGTGAGTAC GTTTTCAG CCGTCTGACT GACACTTGA
 160 A V V I A F O D G A C M A L I S L H I F Y K K C A O L T V N I

30/35

FIGURE 8B

801 TACTGCGTG AGGATGGCCA GTGGGCGGAA CAGCGGCTCA (GGCTGTCAG CTGTGCTCGG GAGTTGAGG CAGCTGAGGG GAAACCAAG TGGCGAGCTT
 ATGAGAGCAC TCTTACCGGT CACCGAGCTT GTTGTTCATT TCTGAGCTC GACAGGAGTC CCCAGCTCC GTGACTCC CTGTGCTTC ACGCTCGGA
 210 Y C R E D G Q W A E Q P V T G C S C A P G F E A A E G N T K C R A C

 901 GTGCCAGGG CAGCTTCAG CCCCTTCAG GATAGCTC GTGCCAGCA TGCCAGCCA ATAGCACTC TAACCACTT GATCAGCGG TCTCCAGTG
 CAGGCTCC GTGGAGTTC GGGCAATTC CTATTTCAG GAGGTGCTT AGGTGCTGT TATGCTGAG ATTGTGTAA CTTAGTCCG AGAGCTCAC
 272 A Q G T P K P L S G E G S C O P C P A N S H E N T I G S A V C Q C

 1001 CCGGTCGG TACTTCGGG CAGGACAGA CCCCGGCTT GCACCTTCA CCACCTCTCC TTGCGCTCG CCGAGCTGG TTTCCTCTT GAACTCTCC
 GAGGAGCCC ATGAGGCCC GTGCTGTCT GAGGCTCTT GAGGCTCTT CTTGAGCTT GTTGGAGG AGCGGAGG GCTTGCACC AAGGCGGA CTGCGCGAGG
 305 R V G Y F R A R T D P R G A P C T T P P S A P R S V V S M L N G S

 1101 TCTCTGACC TGGATGGG TGCCCGCTG GATCTGTG GCGAGAGGA CTTCACTTAC GCTCTCTCT CCGCTCTCT CCGAGCTGG CCGACCCGGA GCTCTCTG
 AGGAGCTGG ACTTACTC ACGGAGAC CTCAGACAC GGTCTCTCT GAGTGTAT CCGAGCTGG CCGCTCTAC CCGCTCTCT CCGAGCTGG CCGAGCTGG
 318 S L H L E W S A P L F S G G R E D L T Y A L R C N E C R P G G S C A

 1201 CGCCTGCG GAGAGACTG ACTTTGACC CCGGCGCTG GAGCTGCTG GAGCTGCTG TGCTGCTGG AGGCTACTT CTTGACTTCA CTTATACCTT
 GCGGAGCCC CCGCTGAG TGAAGCTG GCGGCGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG
 372 P C G G D L T F D P G P R D L V E P M V V V R G L R P D P T Y T P

 1301 TGAGTCACT GATTTGAGG GGTATCTTC CTTAGCTG GCGGCGCTG CATTGAGC TGTCACTC ACCACTGACC GAGAGTACC TCTCTGAGT
 ACTCAGTCA GTTACTTCC CCGATAGG GATCTGCT GATCTGCT CCGGCGCTG GTAACTGG ACAGTACAG TGTGACTGG CTCTCATGG AGGACTCAC
 405 E V T A L N G V S S L A T G P V P F E P V N V T T D R E V P P A V

 1401 TCTGATCC GGTGAGCG GTCTCACC AGGACTTCA GCTGCTG GCTGCTG CCGGACCCA GTGCTGCT GCTGCTGCT GCTGCTGCT GCTGCTGCT
 AGGCTGAG CCGACTGCG CAGGCTGCT TGTGCTGCT CTTGAGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG CTTGAGCTG
 418 S D I R V T R S S P S S L S L A M A V P R A P S G A V L D Y E V R Y

31/35

FIGURE 8C

1501 AGATGAGAA GGGGCGGAG GGTCCGAGTA GGTTCGGTCTT CTTGAGAGG TCAGAAACC GGGGAGACT GGGGGGCTG AAGGGGGAG CCACTACTT
 TGTACTCTT CCGCGGCTC CCAGGCTTCTT CAGAGCTTCTT AGTCTTCTT CCGCTCTGTA CCGGGGCGAC TTGGGGGCTC GGTGGATGA
 472 H E R G A E G P S S V R F I K T S E N R A E L R G L K R O A S Y L

 1601 GGTTCAGCTA CCGGGGCGCT CTGAGGCGG CTATGAGGCTT TCGGCGAGG AACATCAGG CCAGACCAA CTGATGAGA GCGAGGCTG GCGGAGCAG
 CCAGTCCAT GCGGCGGCA GACTCGGCGG GATGCTGAGG AATCGGCTC TTGATGTC GGTCTGGT GACTACTCT CCGTCCGAG CCGCTCTGTC
 505 V Q V R A R S E A G Y G P F G Q E H S Q T O L D E S E O W R E Q

 1701 CTGCGCTTA TTGGGCGAC GCGAGTCTG GGTTCGCTC GGTTCGCTT GGTCTCTG GTCGAGTTC TCTGCTCAG GAGGAGAGG AATCGAGAG
 GACCGGACT AACGCGGCT CCGTCAGC CCAGCAGG ACAGAGGCA CCAGTAACAC CAGCTCAAG AGCGGAGTC CTTCCTCTCO TTACCTCTC
 538 L A L I A G T A V V G V V L V L V V L V V A V L C L R R Q S M G R E

 1801 AAGCAGATA TTGAGCAA CAGGAGAGT ATCTCATGG ACATGGTACT AAGTCTACA TCGAGGCTT CACTTATGA GACCTTAATG AGGCTGTGAG
 TTGCTTAT AAGCTGTTT GTGCTGTCA TAGAGTAGC TTATACATGA TTCCAGATGT AGCTGGGAA GTGANTACTT CTGGATTAAC TCGGACACTC
 572 A E Y S D K H G Q Y L I G H G T K V Y I D P F T Y E D P N E A V R

 1901 GGAATTGCA AAGAGATCG ATGTCTCTTA CGTCAGATT GAGAGGTGA TTGTGCGAG TTGTTTGGC GAGTGTGCC GGGGGGCTT CAAAGGCGCA
 CCTTAACGT TTCTCTAGC TACAGAGGAT GCAGTCTTA CTTCCTCCTT AACGAGTCC ACTCAAGCG CTCCAGAGG CCGGGGCGTA GTTCGCGGT
 605 E F A K E I D V S V V K I E E V I G A G E F G E V C R G R L K A P

 2001 GCGAAGAG AGAGTGTGT GCGATCAAG ACCGTGAAG GTTCTACAC GAGGCGGAG CCGGTGAGT TTCTGAGCA GCGCTCCATC ATGGGCGCT
 CCGTCTTCC TCTGAGACA CCGTTAGTTC TCGACTTTC CACCATGTC CCGCGGTC CTTCGCTC AGACTCTCT CCGAGGATAG TACCGGCTCA
 638 G K K E S C V A I K T L K G G Y T F R Q R R E F L S E A S I M G Q P

 2101 TCGAGACCC CAATATCAT CCGCTGAGG GGTGCTAC CAGAGCATG CCGGTATGA TTCTACAGA GTTCTATGAG AACGGGCGCC TCGACTCTT
 AGCTGCTG GATTATAGT GCGGACCTC CCGACCATG GTTGTCTAC CCGCAGTCT AGAGTGTCT CAGTACTCT TTGCGGCGG ACCTGAGGA
 672 E H P N I I R L E G V V T N S M P V M I L T E P M E M G A L D S F

FIGURE 8D

2201 CTTGCGGCTA AACGACGAC AGTTACAGT CATCCAGTC GTGGGATCC TCGGGGAT CCGCTGGGC ATCGGTACC TTGCGAGAT GAGCTAGTC
 GACGCGCAT TTGCGCTG TCAAGTCA GTAGGTGAG CACCGTAGC AGCGCGTA GCGAGCGCG TACCGATG AACCGCTTA CTGATGAG
 705 I R L N D G Q F T V I O L V G M L R G I A S G M R Y L A E M S Y V

2301 CACCGAGCC TGGCTGCTG CACATCTTA GTCAACACA ACTGCTGT CANAGTGT GACTTGACC TTTCGGAAT CCTGAGGAG AACTCTTCG
 GTGCTCTGG ACCGCGAGC GTTGAGGAT CATTGTGT TGGAGGAG GTTACAGAG CTGAACCGG AAGCGCTAA GGAAGCTCTC TTGAGAGGC
 738 H R D L A A R N I L V N S N L V C K V S D P G L S R F L E E N S S D

2401 ATCCACCTA CACGAGCTC CTGGAGGAA AGATTCCAT CCGATGACT GCGCGGAG CCAATGCTT CCGGAGTTC ACTTCGCGA GTGATGCTG
 TAGH:GGAT GTGCTGAGG GACCTGCTT TCTAAGGTA GCTACCTGA CCGGCTCC GGTAAAGGAA GCGCTTCAAG TGAAGCGGT CACTACGAGC
 772 P T Y T S S L G Q K I P I R W T A P E A I A F R K F T S A S D A W

2501 GAGTTACGG ATTGTGAT GTGAGTGAT GTGCTTGG GAGGCGCT ACTGGACAT GAGCAATCAG GACGTGATCA ATGCAATTA ACAGACTAC
 CTCATGCCC TAACACTACA CCGTCCACTA CATTAAACCC CTCGCGCA TACCTGTA CTGCTAGTC CTGCACTAGT TAGGTAACT TGCTCTGATG
 805 S Y G I V M W E V M S F G E R P Y W D M S N Q D V I M A I E Q D Y

2601 CAGCTGCCC CCGCCCGA CTGTCGCC TCGTCCACC AGTCATGCT GACTGTGG CAGAAAGACC GGAATGCGG GCGCGCTTC CCGAGTGG
 GCGACCGG GCGGCTCT GACAGGTGG AGGAGTGG TCGAGTACA CTTGACACC GTCTTCTGG CTTACCGGC CCGGCGGAG GCGTCCACC
 838 R L P P P P D C P T S L H Q L M L D C W Q K D R M A R P R F P Q V V

2701 TCAGCGCCT GAGACAGAT ATCGGAAC CCGCAGCT CAATGCTG GCGCGGAGA ATGCGGCGC CTCACACCT CTCTGAGCC AGCGAGCC
 AGTCGCGA CTTGCTTAC TAGCGCTGG GCGGTGCGA GTTTAGAC CCGGCTCT TACCGCGG GAGTGTGGA GAGGAGCTG TCGCGTGG
 872 S A L D R M I R N P A S L K I V A R E N G G A S H P L L D Q R Q P

2801 TCACTACTA GCTTTGCT CTGCGGGA GTGCTTGG GCAATAAA TGGAGATA CAGAAAGT TTGCGAGCC CTGCTTGG CTGCTGAG
 ACTATGACT CGAAACCA GACAGCTCT CACGAAACCG CGTAGTTT ACCTTCTAT ACCTTCTA AGCGTGGC GACCGAAC CAGGAGCTC
 905 H Y S A F G S V G E W L R A I K M G R Y E E S F A A G F G S F E

FIGURE 8E

2901 CTGCTGAGCC AGATCTCTGC TGAGCACTG CTGCAATCG GATCACTCT GGGGGGACAC CAGAGAAA TCTTGGCCAG TGTCCAGCAC ATCAAGTCCC
 CACCAGTCGG TCTAGAGAG ACTCTTACG GAKCTTACG CTCAGTGAGA CCGCTCTG GTCTCTTT AGAACGGTC ACAGTCTGT TACTTCAGCG
 918 L V S Q I S A E D I L R I G V T L A G H Q R K I L A S V Q H M K S Q

1001 AGGCCAAGCC GGAAGCCCG GRTGACAG GAGGACCGC CCGCAGTAC TGACCTGAG GACTCCCA CCCCAGGAC ACCGCTCC CATTTCCCG
 TCGGTTCCG CCGTTGCG CACCTGTC CTCTGCGG GGGCTCAT ACTGACGTC CTGAGGGT GGGTCCCTG TGGGAGGG GTAAAGGCC
 972 A K P G T P G G T G G P A P Q Y O P A O T P H P R D T A S P F S Q

1101 GGCAGAGTGG GGACTCAGAG AGGCCCCAG CCGTGTCCC CCGTGATTG CACTTGAGC CCGTGGGTTG AGAGTTGCG AATTGGAGA GACAGATTT
 CCGTCTACC CCGTCTGTC TCGGGGTC GGCACAGCG GGACTTAC GTTAACTG GGCACCCAC TCTCAACCG TTAACTCT CTGTCTAA
 1005 A E W G L T E A P S P V P R W I A L O A R Q V R S W Q F G E T G F

1201 GGGGTTCTG CCATAATGG AGGGGAGAT CACCCCGAG CCACTCGG GAACTCAGA CCAAGGTTA GGGGCTTT CCGTCAGAC TGGTCTGAC
 CCCCAGAC GGTATTATC TCCCTTTTA GTGGGGTC GGTGAGCC CTRAGGCT GTTCCACT CCGCGGAA GCGAGTCTG ACCCAGCTG
 1018 G G S A I I G G E N H P P A T S G N S R P R V R A P F P Q D W V O P

1301 CAGAGAAA GGAAGTCCC ACATCTCC AGCTCCCA GTGGCCCC TCACCTTAT GGTGCTTC CCGAGACA AGAGAGTGT GACTCCCTG
 GTCTCTTT CTTACAGG TTGTAGAGG TCGAGGGT CCAAGGGG AGTGAACTA CCGAGGAG GCGTCTGT TTCTCACA CTAGGAGAC
 1072 E E K E V P N I S Q P P Q V P P S P O M V R S R R P K R V O L P C

1401 CCAGTCCAG AGTGGGGG CTGTCCAG GGCAGAGAG GGGTGCAG CCGCAGTAC AAATCATTT GGGTTGTAG TCCCACTTG CTGCTGTCAC
 GGTTCAGTC TCACCCCC GACAGGTC CCGTTCTTC CCGCAGTCC CCGTCACTG TTTAGTAC CCGAACATC AGGTGAGAC GAGCAGCTG
 1105 Q L Q S G G A V P G G K K G C Q G P V T K S L G F V V P T C C C H

1501 CACCAACTC ATCATTTTT TTCCCTTGA ATGCCCTC CCGCAGTCC TCGCTTATA TTGAGTTT TCGAGTTT TTTTGTCT TAATTTTCT
 GTGTTTGG TTAGTAANA AGGACAT TTAGGTGAG GGGTGGAG ACCGAGTAT AACTTCANA AACTCAAC AANAACAGA ATTAAAGA
 1118 U Q T Q S P F S L V N A P P P A A A F I L R V F E P C F W S O F F S

34/35

FIGURE 8F

3601 CCGCGTCCC TTTTGTTC TCGTTTGT TTTCTATCG TCCTGTGCAT AACTTGTGT TCGAGGAGC CTGTTTCACT ATGCGCTCT TTGCCCCAGT
 GCGCGAAGG AANAACAAAG AAGCAACACA AANAGATGCG AGGAACAGTA TTGAACACA ACCTGCTTG GACAAATGA TACCGAGCA ACGGTTCA
 1172 P F P F C F F V L F F V R P C H N F V L E G T C F T M A S P A Q V

3701 TGAACAGGG GCCATCATC ATGCTCTTT CCAGAACAGT GCTTGTCA TCCCATCC CCGAGCCCC CCGAGCTGG CCAGCTGG TCCTATGAAG
 ACTTGTCCC CGGTAGTAG TACAGACAAA GGTTTGTCA CGAACAGT AGGTGTAG GGCCTGGG GACCTGGG GTTGGACAG AGGATACTTC
 1205 E T G A H H V C F Q N S A L V I P H P R T P P G T P K L C P H K

3801 GCGTGTGGG TGAGGTAGT AAGCGCGG TAGTTGTGG TGAACCCAG AAGCGAGCC CGTGTCTTG AGGCTTCTT AATTATATT TAAAGTGA
 CCACACCCC ACTCCATCAG TTTCCCGCC ATCAACCACC ACTTGGTC TTTGCTGG GCCACGACC TCCCAAGAA TTTATATAA ATTTTTCAT
 1218 G C G V R O O K G R O L V V E P R N G R R C L E G F L N Y I O K S N

3901 ACTTTTGT TAAATAAAG AATGACAC GTTCCAGC TCGAGGGA AANAANA AANAANA
 TGAACAT ATTTATTC TTTTACCTG CACAGGCTG AGTCCCAT TTTTTTT TTTTTT
 1272 F L Y K O K K M G R V P A P G V K K K K R K

ARNILVNSNLVCKVSDFGLSRFLEDDTSDPTTYSALGGKIPHRHTAPEAIQYRKFFASAS

FIGURE 9
bptk 1

NVLKSPNHVKITDFGLARLLEGEDEKEYNADGGKMPKWNHLECIHYRKFTTHQS

FIGURE 10
bptk 2

NCHLAGDMTVCVADFGLSWKIYSGATIVRGCAKLPVKNLALGSLADNLYTVHS

FIGURE 11
bptk 3

NCLVGKNYTIKIADFGMSRNLYSGDY

FIGURE 12
bptk 4

TRNILVENENRVKIGDFGLTKVLPQDKEYYKVKEPGESPIFYAPESLTESLFSVASD

FIGURE 13
bptk 5

ARNILVNSNLVCKVSDFGMSRVLEDDPEAAVYTRGGKIPIRHTAPEAIYRKFTSASD

FIGURE 14
bptk 7

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 93/00586

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5 C12N15/12; C12N15/54; C12N9/12; //C12Q1/68, C12N15/11		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
Int.Cl. 5	C12N ; C12Q ; C07K	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are indicated in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	<p>NEURON vol. 6, no. 5, May 1991, pages 691 - 704 LAI, C. & LEMKE, G. 'An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system' see the whole document</p> <p style="text-align: center;">---</p> <p style="text-align: right;">-/--</p>	1-7
<p>¹⁰ Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
07 JUNE 1993		02 -07- 1993
International Searching Authority		Signature of Authorized Officer
EUROPEAN PATENT OFFICE		ANDRES S.M.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

OBSCURITIES

Claims 2,3,5 and 7 as filed in the application are in contradiction with the description. Indeed, they are dependent on claims 1,4 or 6 which are related to DNAs or proteins of human MEGAKARYOCYTIC origin, whereas they claim some sequences which are not from such an origin, if taking reference to page 8 of the description. Therefore the search of these claims has been made independently of claims 1,4 and 6.

Claims 14 and 15 relating to the nucleotide and the protein sequences of SAL-D4 are filed as being dependent on claim 10 which relates to SAL-S1. They have been interpreted and searched as depending on claim 13.

Claims 17 and 18 relating to the nucleotide and the protein sequences of LpTK 3 are filed as being dependent on claim 14 which relates to SAL-D4. They have been interpreted and searched as depending on claim 16.

Claims 20 and 21 relating to a DNA expression vector and a cell transformed therewith, are filed as being dependent on claim 17 which relates to a protein (LpTK 3). They have been searched as depending on claim 19.

Finally, claim 2, point a) relating to SAL-S1 refers erroneously to SEQ ID 7 which describes SAL-D4. It has been interpreted as referring on SEQ ID 5 and 17).

US 9300586
SA 69794

07/06/93

EPO FORM 1079

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

II. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 88, no. 23, 1 December 1991, WASHINGTON US pages 10411 - 10415 HOLTRICH, U. ET AL. 'Two additional protein-tyrosine kinases expressed in human lung: fourth member of the fibroblast growth factor receptor family and an intracellular protein-tyrosine kinase'</p> <p>---</p>	1-7, 13-15
A	<p>GENE vol. 110, no. 2, 15 January 1992, AMSTERDAM NL pages 205 - 211 BRXUNINGER, A. ET AL. 'Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases' see the whole document</p> <p>---</p>	1-7, 13-15
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 87, November 1990, WASHINGTON US pages 8913 - 8917 PARTANEN, J. ET AL. 'Putative tyrosine kinases expressed in K-562 human leukemia cells' see especially clone JTK10</p> <p>---</p>	1-7
A	<p>MOLECULAR AND CELLULAR BIOLOGY vol. 9, no. 4, April 1989, WASHINGTON US pages 1587 - 1593 HAO, Q.-L. ET AL. 'Isolation and sequence analysis of a novel human tyrosine kinase gene' see the whole document</p> <p>---</p>	1-7, 16-18
P,A	<p>WO,A,9 214 748 (AMERICAN CYANAMID COMPANY) 3 September 1992 see the whole document</p> <p>-----</p>	1-12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 93/00586

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
obscurities.
see additional page

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.